```
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 16 AA;

  Query Match             81.9%;  Score 68;  DB 5;  Length 16;
  Best Local Similarity   87.5%;  Pred. No. 1.8e-05;
  Matches   14;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              |||| |||||||| ||
Db          1 SVIATQMTYKVYMPGT 16


RESULT 8
ABB99071
ID   ABB99071 standard; peptide; 16 AA.
XX
AC   ABB99071;
XX
DT   22-JAN-2003  (first entry)
XX
DE   N-terminal amino acid sequence of a CFM #11.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 4; Page 281; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 16 AA;

  Query Match             80.7%;  Score 67;  DB 5;  Length 16;
  Best Local Similarity   87.5%;  Pred. No. 2.8e-05;
  Matches   14;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              || | ||||||||||
Db          1 SVSATQMTYKVYMSGT 16


RESULT 9
ABB99069
ID   ABB99069 standard; peptide; 16 AA.
XX
AC   ABB99069;
XX
DT   22-JAN-2003  (first entry)
XX
DE   N-terminal amino acid sequence of a CFM #9.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 4; Page 280; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
```

```
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 16 AA;

  Query Match             79.5%;  Score 66;  DB 5;  Length 16;
  Best Local Similarity   87.5%;  Pred. No. 4.3e-05;
  Matches   14;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              | || ||||||||||||
Db          1 SGIATQMTYKVYMSGT 16


RESULT 10
ABB99074
ID   ABB99074 standard; peptide; 16 AA.
XX
AC   ABB99074;
XX
DT   22-JAN-2003  (first entry)
XX
DE   N-terminal amino acid sequence of a CFM #14.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 10
FT                   /label= Xaa
FT                   /note= "Xaa is any amino acid except Lys"
FT   Misc-difference 11
FT                   /label= Xaa
FT                   /note= "Xaa is any amino acid except Val"
FT   Misc-difference 13
FT                   /label= Xaa
FT                   /note= "Xaa is any amino acid except Met"
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 4; Page 282; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 16 AA;

  Query Match             74.7%;  Score 62;  DB 5;  Length 16;
  Best Local Similarity   81.2%;  Pred. No. 0.00023;
  Matches   13;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              |||||||||  | |||
Db          1 SVIAKQMTYXXYXSGT 16


RESULT 11
ABP70008
ID   ABP70008 standard; peptide; 13 AA.
XX
AC   ABP70008;
XX
DT   06-AUG-2003  (revised)
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 184.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Pavona decussata.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
```

```
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 473; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 13 AA;

  Query Match             43.4%;  Score 36;  DB 5;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMT 8
              ||||||||
Db          1 SVIAKQMT 8


RESULT 12
AAR77526
ID   AAR77526 standard; peptide; 14 AA.
XX
AC   AAR77526;
XX
DT   25-MAR-2003  (revised)
DT   12-JUN-1996  (first entry)
XX
DE   p45 metalloprotease N-terminal fragment.
XX
KW   Metalloprotease; enzyme; MP; p45; fusarium oxysporum; bacillus;
KW   thermolysin; casein; Aspergillus oryzae.
XX
OS   Fusarium oxysporum.
XX
PN   WO9530757-A2.
XX
PD   16-NOV-1995.
XX
PF   03-MAY-1995;   95WO-US005534.
XX
PR   04-MAY-1994;   94US-00238108.
PR   03-MAR-1995;   95US-00398489.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
PA   (NOVO ) NOVO-NORDISK AS.
XX
PI   Shuster JR,  Moyer DL,  Madden M,  Fuglsang C,  Branner S;
XX
DR   WPI; 1995-404122/51.
XX
PT   Fungal metallo:protease converts pro:enzyme to active form - has
PT   thermolysin-like activity, useful to cleave pro-sequence of pro:enzyme to
PT   generate mature enzyme.
XX
PS   Claim 12; Page 36; 62pp; English.
XX
CC   AAR77525-R77527 represent the N-terminal sequences of a fungal
CC   metalloprotease (MP). This sequence represents the N-terminus of Fusarium
CC   oxysporum MP p45 (see AAR77528). AAR77525 represents the consensus N-
CC   terminal sequence of the MP from F.oxysporum and Aspergillus oryzae. p45
CC   is a new MP, and has 10 times more efficiency than Bacillus MP. Bacillus
CC   MP is more effective in cleaving primary amino groups from casein. p45
CC   has thermolysin-like activity, and is used to cleave a pro-sequence from
CC   a recombinant proenzyme to generate an active mature enzyme. The MP may
CC   be added to, or produced in, the broth where the proenzyme is being
CC   formed by a recombinant host cell converted with a vector containing the
CC   DNA encoding p45. The MP can also be used to assay the level of
CC   activatable proenzyme in a sample. (Updated on 25-MAR-2003 to correct PA
CC   field.)
XX
SQ   Sequence 14 AA;

  Query Match             34.9%;  Score 29;  DB 2;  Length 14;
  Best Local Similarity   75.0%;  Pred. No. 2.2e+02;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          8 TYKVYMSG 15
              |||||  |
Db          2 TYKVYPWG 9


RESULT 13
AAW05846
ID   AAW05846 standard; peptide; 14 AA.
XX
AC   AAW05846;
XX
DT   16-OCT-2003  (revised)
DT   28-JAN-1997  (first entry)
XX
DE   Fusarium oxysporum p45 metalloprotease N-terminal peptide.
XX
KW   Metalloprotease; protease; p45; recombinant protein; host cell.
XX
OS   Fusarium oxysporum; strain DSM 2672.
XX
PN   WO9629391-A1.
XX
PD   26-SEP-1996.
XX
PF   20-MAR-1996;   96WO-DK000111.
XX
PR   20-MAR-1995;   95DK-00000284.
XX
PA   (NOVO ) NOVO-NORDISK AS.
XX
PI   Lehmbeck J;
XX
DR   WPI; 1996-443168/44.
XX
PT   Host cell with reduced expression of metallo-protease - for prodn. of
PT   recombinant proteins, opt. as their precursors.
XX
PS   Example 1; Page 34; 51pp; English.
```

```
XX
CC   The N-terminal sequence (AAW05846) of Fusarium oxysporum DSM 2672 p45
CC   metalloprotease (see also AAW05845) was identified by amino acid analysis
CC   of a protein isolated from a fermentation broth. A PCR primer based on
CC   this peptide was used, together with a primer based on a p45 internal
CC   peptide, in the PCR cloning of the p45 gene (AAT40133) from F. oxysporum
CC   genomic DNA. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 14 AA;

  Query Match             34.9%;  Score 29;  DB 2;  Length 14;
  Best Local Similarity   75.0%;  Pred. No. 2.2e+02;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          8 TYKVYMSG 15
              |||||  |
Db          2 TYKVYPWG 9


RESULT 14
AAM48968
ID   AAM48968 standard; peptide; 15 AA.
XX
AC   AAM48968;
XX
DT   25-APR-2002  (first entry)
XX
DE   Human zinc finger protein 53 N-terminal peptide.
XX
KW   Human; zinc finger protein 53; cancer; nervous system disease;
KW   development disorder; metabolic disease; inflammation; haemopathy;
KW   immunological disease; HIV infection; gene therapy.
XX
OS   Homo sapiens.
XX
PN   CN1314368-A.
XX
PD   26-SEP-2001.
XX
PF   17-MAR-2000; 2000CN-00114979.
XX
PR   17-MAR-2000; 2000CN-00114979.
XX
PA   (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX
PI   Mao Y,  Xie Y;
XX
DR   WPI; 2002-056224/08.
XX
PT   New polypeptide-human zinc finger protein 53 and polynucleotide for
PT   coding such polypeptide.
XX
PS   Example 6; Page 18(Disclosure); 33pp; Chinese.
XX
CC   The present invention provides the protein and coding sequences of human
CC   zinc finger protein 53. The sequences can be used in the treatment of
CC   cancer, haemopathy, nervous system disorders, development disorders,
CC   metabolic disorders, inflammation, immunological diseases and HIV
CC   infection. The present sequence is the N-terminus of the protein of the
CC   invention
XX
SQ   Sequence 15 AA;

  Query Match             34.9%;  Score 29;  DB 5;  Length 15;
  Best Local Similarity   54.5%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          5 KQMTYKVYMSG 15
              | || | : ||
Db          2 KNMTLKSFASG 12


RESULT 15
AAW39598
ID   AAW39598 standard; peptide; 11 AA.
XX
AC   AAW39598;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human melanoma associated protein tyrosinase peptide (pos. 367-377).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
PF   28-APR-1997;   97WO-NL000229.
XX
PR   26-APR-1996;   96EP-00201145.
PR   23-DEC-1996;   96EP-00203670.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI   Van Der Burg SH,  Kast WM,  Toes REM,  Offringa R,  Melief CJM;
XX
DR   WPI; 1997-549891/50.
XX
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells.
XX
PS   Example 3; Page 75; 109pp; English.
XX
CC   Peptides AAW39430-W39734 are used in a novel method for the selection of
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC   method involves the identification of peptide sequences capable of
CC   binding to an HLA (human leukocyte antigen) class I molecule and
CC   measuring the binding of this epitope peptide to the HLA class I peptide.
CC   The stability of binding of the peptide and MHC (major histocompatablilty
CC   complex) class I molecule is measured on intact human B cells carrying
CC   the MHC molecule at their cell surfaces. The method can be used to select
CC   peptide epitopes for generating vaccines against a disease associated
CC   with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC   especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC   immune responses. Peptide AAW39598 is derived from the human melanoma
CC   associated protein tyrosinase which is capable of upregulating HLA-A*0201
CC   molecules on T2 cells
XX
SQ   Sequence 11 AA;

  Query Match             33.7%;  Score 28;  DB 2;  Length 11;
  Best Local Similarity   57.1%;  Pred. No. 2.5e+02;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          11 VYMSGTV 17
               :||:||:
Db           2 IYMNGTM 8


Search completed: August 12, 2004, 07:03:22
Job time : 51 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 07:04:56 ; Search time 41 Seconds
(without alignments)
130.165 Million cell updates/sec

Title: US-09-890-463-2
Perfect score: 83
Sequence: 1 SVIAKQMTYKVYMSGTV 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 228781

Minimum DB seq length: 0
Maximum DB seq length: 17

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	30	36.1	11	15	US-10-360-101-187	Sequence 187, App
2	28	33.7	9	12	US-10-253-286-492	Sequence 492, App
3	28	33.7	9	12	US-10-253-286-493	Sequence 493, App
4	28	33.7	9	15	US-10-245-871-492	Sequence 492, App
5	28	33.7	9	15	US-10-245-871-493	Sequence 493, App
6	28	33.7	14	12	US-10-253-286-508	Sequence 508, App
7	28	33.7	14	12	US-10-253-286-509	Sequence 509, App
8	28	33.7	14	15	US-10-245-871-508	Sequence 508, App
9	28	33.7	14	15	US-10-245-871-509	Sequence 509, App
10	28	33.7	17	12	US-10-253-286-501	Sequence 501, App
11	28	33.7	17	12	US-10-253-286-510	Sequence 510, App
12	28	33.7	17	15	US-10-245-871-501	Sequence 501, App
13	28	33.7	17	15	US-10-245-871-510	Sequence 510, App
14	27	32.5	10	14	US-10-200-708-291	Sequence 291, App
15	27	32.5	13	14	US-10-246-581-24	Sequence 24, Appl
16	26	31.3	10	8	US-08-325-278-12	Sequence 12, Appl
17	26	31.3	10	12	US-10-416-822-4	Sequence 4, Appli
18	26	31.3	12	14	US-10-246-581-22	Sequence 22, Appl
19	26	31.3	13	12	US-09-988-493-253	Sequence 253, App
20	26	31.3	14	16	US-10-762-629-48	Sequence 48, Appl
21	26	31.3	16	14	US-10-225-567A-2292	Sequence 2292, Ap
22	26	31.3	17	11	US-09-754-831A-34	Sequence 34, Appl
23	25	30.1	8	15	US-10-388-337-11	Sequence 11, Appl
24	25	30.1	9	9	US-09-812-528-5	Sequence 5, Appli
25	25	30.1	9	9	US-09-847-185-38	Sequence 38, Appl
26	25	30.1	9	9	US-09-923-831-21	Sequence 21, Appl
27	25	30.1	9	9	US-09-872-832-13	Sequence 13, Appl
28	25	30.1	9	9	US-09-888-721-27	Sequence 27, Appl
29	25	30.1	9	9	US-09-888-721-28	Sequence 28, Appl
30	25	30.1	9	9	US-09-766-889A-33	Sequence 33, Appl
31	25	30.1	9	9	US-09-909-460-60	Sequence 60, Appl
32	25	30.1	9	10	US-09-898-860-32	Sequence 32, Appl
33	25	30.1	9	11	US-09-077-439A-7	Sequence 7, Appli
34	25	30.1	9	12	US-10-218-095-28	Sequence 28, Appl
35	25	30.1	9	12	US-10-253-286-519	Sequence 519, App
36	25	30.1	9	12	US-09-775-805-12	Sequence 12, Appl
37	25	30.1	9	12	US-09-077-214-11	Sequence 11, Appl
38	25	30.1	9	12	US-10-289-566-1	Sequence 1, Appli
39	25	30.1	9	12	US-10-367-580-145	Sequence 145, App
40	25	30.1	9	12	US-10-367-593-145	Sequence 145, App
41	25	30.1	9	12	US-10-367-594-145	Sequence 145, App
42	25	30.1	9	12	US-10-367-654-145	Sequence 145, App
43	25	30.1	9	12	US-10-367-658-145	Sequence 145, App
44	25	30.1	9	12	US-10-367-668-145	Sequence 145, App
45	25	30.1	9	12	US-09-872-836-60	Sequence 60, Appl

ALIGNMENTS

RESULT 1
US-10-360-101-187
; Sequence 187, Application US/10360101
; Publication No. US20040009550A1
; GENERAL INFORMATION:
; APPLICANT: Moll, Gert N.
; APPLICANT: Leenhouts, Cornelis J.
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
; FILE REFERENCE: 2183-5673
; CURRENT APPLICATION NUMBER: US/10/360,101
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: EP 02077060.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 309
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 187
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence of interleukin-2 fragment (60-70) (T-cell growth factor)
US-10-360-101-187

Query Match 36.1%; Score 30; DB 15; Length 11;
Best Local Similarity 62.5%; Pred. No. 89;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy          7 MTYKVYMS 14
              :|:| |||
Db          1 LTFKFYMS 8
```

RESULT 2
US-10-253-286-492
; Sequence 492, Application US/10253286
; Publication No. US20040058881A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT

```
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
;  FILE REFERENCE: REH-2015
;  CURRENT APPLICATION NUMBER: US/10/253,286
;  CURRENT FILING DATE:  2003-01-13
;  PRIOR APPLICATION NUMBER: 10/197,000
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: 09/396,813
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 905
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 492
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-253-286-492

  Query Match             33.7%;  Score 28;  DB 12;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 1.2e+06;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 VYMSGTV 17
              :||:||:
Db          3 IYMNGTM 9


RESULT 3
US-10-253-286-493
; Sequence 493, Application US/10253286
; Publication No. US20040058881A1
; GENERAL INFORMATION:
;  APPLICANT: HUMPHREYS, ROBERT
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
;  FILE REFERENCE: REH-2015
;  CURRENT APPLICATION NUMBER: US/10/253,286
;  CURRENT FILING DATE:  2003-01-13
;  PRIOR APPLICATION NUMBER: 10/197,000
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: 09/396,813
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 905
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 493
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-253-286-493

  Query Match             33.7%;  Score 28;  DB 12;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 1.2e+06;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 VYMSGTV 17
              :||:||:
Db          1 IYMNGTM 7


RESULT 4
US-10-245-871-492
; Sequence 492, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
;  APPLICANT: HUMPHREYS, ROBERT
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
;  FILE REFERENCE: REH-2013
;  CURRENT APPLICATION NUMBER: US/10/245,871
;  CURRENT FILING DATE:  2003-01-09
;  PRIOR APPLICATION NUMBER: 10/197,000
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: 09/396,813
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 905
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 492
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-245-871-492

  Query Match             33.7%;  Score 28;  DB 15;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 1.2e+06;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 VYMSGTV 17
              :||:||:
Db          3 IYMNGTM 9


RESULT 5
US-10-245-871-493
; Sequence 493, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
;  APPLICANT: HUMPHREYS, ROBERT
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
;  FILE REFERENCE: REH-2013
;  CURRENT APPLICATION NUMBER: US/10/245,871
;  CURRENT FILING DATE:  2003-01-09
;  PRIOR APPLICATION NUMBER: 10/197,000
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: 09/396,813
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 905
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 493
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-245-871-493

  Query Match             33.7%;  Score 28;  DB 15;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 1.2e+06;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 VYMSGTV 17
              :||:||:
Db          1 IYMNGTM 7


RESULT 6
US-10-253-286-508
; Sequence 508, Application US/10253286
; Publication No. US20040058881A1
; GENERAL INFORMATION:
;  APPLICANT: HUMPHREYS, ROBERT
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
;  FILE REFERENCE: REH-2015
;  CURRENT APPLICATION NUMBER: US/10/253,286
;  CURRENT FILING DATE:  2003-01-13
;  PRIOR APPLICATION NUMBER: 10/197,000
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: 09/396,813
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 905
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 508
;   LENGTH: 14
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Ii-key/tyrosinase overlapping hybrid peptide
```

```
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (5)
;   OTHER INFORMATION: a-aminovaleric acid
;   FEATURE:
;   OTHER INFORMATION: C-term amidated
US-10-253-286-508

  Query Match             33.7%;  Score 28;  DB 12;  Length 14;
  Best Local Similarity   57.1%;  Pred. No. 2.7e+02;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 VYMSGTV 17
              :||:||:
Db          8 IYMNGTM 14


RESULT 7
US-10-253-286-509
; Sequence 509, Application US/10253286
; Publication No. US20040058881A1
; GENERAL INFORMATION:
;  APPLICANT: HUMPHREYS, ROBERT
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
;  FILE REFERENCE: REH-2015
;  CURRENT APPLICATION NUMBER: US/10/253,286
;  CURRENT FILING DATE:  2003-01-13
;  PRIOR APPLICATION NUMBER: 10/197,000
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: 09/396,813
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 905
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 509
;   LENGTH: 14
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Ii-key/tyrosinase overlapping hybrid peptide
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (5)
;   OTHER INFORMATION: a-aminovaleric acid
;   FEATURE:
;   OTHER INFORMATION: C-term amidated
US-10-253-286-509

  Query Match             33.7%;  Score 28;  DB 12;  Length 14;
  Best Local Similarity   57.1%;  Pred. No. 2.7e+02;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 VYMSGTV 17
              :||:||:
Db          6 IYMNGTM 12


RESULT 8
US-10-245-871-508
; Sequence 508, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
;  APPLICANT: HUMPHREYS, ROBERT
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
;  FILE REFERENCE: REH-2013
;  CURRENT APPLICATION NUMBER: US/10/245,871
;  CURRENT FILING DATE:  2003-01-09
;  PRIOR APPLICATION NUMBER: 10/197,000
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: 09/396,813
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 905
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 508
;   LENGTH: 14
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Ii-key/tyrosinase overlapping hybrid peptide
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (5)
;   OTHER INFORMATION: a-aminovaleric acid
;   FEATURE:
;   OTHER INFORMATION: C-term amidated
US-10-245-871-508

  Query Match             33.7%;  Score 28;  DB 15;  Length 14;
  Best Local Similarity   57.1%;  Pred. No. 2.7e+02;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 VYMSGTV 17
              :||:||:
Db          8 IYMNGTM 14


RESULT 9
US-10-245-871-509
; Sequence 509, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
;  APPLICANT: HUMPHREYS, ROBERT
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
;  FILE REFERENCE: REH-2013
;  CURRENT APPLICATION NUMBER: US/10/245,871
;  CURRENT FILING DATE:  2003-01-09
;  PRIOR APPLICATION NUMBER: 10/197,000
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: 09/396,813
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 905
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 509
;   LENGTH: 14
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Ii-key/tyrosinase overlapping hybrid peptide
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (5)
;   OTHER INFORMATION: a-aminovaleric acid
;   FEATURE:
;   OTHER INFORMATION: C-term amidated
US-10-245-871-509

  Query Match             33.7%;  Score 28;  DB 15;  Length 14;
  Best Local Similarity   57.1%;  Pred. No. 2.7e+02;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 VYMSGTV 17
              :||:||:
Db          6 IYMNGTM 12


RESULT 10
US-10-253-286-501
; Sequence 501, Application US/10253286
; Publication No. US20040058881A1
; GENERAL INFORMATION:
;  APPLICANT: HUMPHREYS, ROBERT
```

```
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
;  FILE REFERENCE: REH-2015
;  CURRENT APPLICATION NUMBER: US/10/253,286
;  CURRENT FILING DATE:  2003-01-13
;  PRIOR APPLICATION NUMBER: 10/197,000
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: 09/396,813
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 905
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 501
;   LENGTH: 17
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-253-286-501

  Query Match             33.7%;  Score 28;  DB 12;  Length 17;
  Best Local Similarity   57.1%;  Pred. No. 3.4e+02;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 VYMSGTV 17
              :||:||:
Db          4 IYMNGTM 10


RESULT 11
US-10-253-286-510
; Sequence 510, Application US/10253286
; Publication No. US20040058881A1
; GENERAL INFORMATION:
;  APPLICANT: HUMPHREYS, ROBERT
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
;  FILE REFERENCE: REH-2015
;  CURRENT APPLICATION NUMBER: US/10/253,286
;  CURRENT FILING DATE:  2003-01-13
;  PRIOR APPLICATION NUMBER: 10/197,000
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: 09/396,813
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 905
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 510
;   LENGTH: 17
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Ii-key/tyrosinase overlapping hybrid peptide
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (5)
;   OTHER INFORMATION: a-aminovaleric acid
;   FEATURE:
;   OTHER INFORMATION: C-term amidated
US-10-253-286-510

  Query Match             33.7%;  Score 28;  DB 12;  Length 17;
  Best Local Similarity   57.1%;  Pred. No. 3.4e+02;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 VYMSGTV 17
              :||:||:
Db          9 IYMNGTM 15


RESULT 12
US-10-245-871-501
; Sequence 501, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
;  APPLICANT: HUMPHREYS, ROBERT
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
;  FILE REFERENCE: REH-2013
;  CURRENT APPLICATION NUMBER: US/10/245,871
;  CURRENT FILING DATE:  2003-01-09
;  PRIOR APPLICATION NUMBER: 10/197,000
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: 09/396,813
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 905
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 501
;   LENGTH: 17
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-245-871-501

  Query Match             33.7%;  Score 28;  DB 15;  Length 17;
  Best Local Similarity   57.1%;  Pred. No. 3.4e+02;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 VYMSGTV 17
              :||:||:
Db          4 IYMNGTM 10


RESULT 13
US-10-245-871-510
; Sequence 510, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
;  APPLICANT: HUMPHREYS, ROBERT
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
;  FILE REFERENCE: REH-2013
;  CURRENT APPLICATION NUMBER: US/10/245,871
;  CURRENT FILING DATE:  2003-01-09
;  PRIOR APPLICATION NUMBER: 10/197,000
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: 09/396,813
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 905
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 510
;   LENGTH: 17
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Ii-key/tyrosinase overlapping hybrid peptide
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (5)
;   OTHER INFORMATION: a-aminovaleric acid
;   FEATURE:
;   OTHER INFORMATION: C-term amidated
US-10-245-871-510

  Query Match             33.7%;  Score 28;  DB 15;  Length 17;
  Best Local Similarity   57.1%;  Pred. No. 3.4e+02;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 VYMSGTV 17
              :||:||:
Db          9 IYMNGTM 15


RESULT 14
US-10-200-708-291
; Sequence 291, Application US/10200708
; Publication No. US20030180314A1
; GENERAL INFORMATION:
;  APPLICANT: DeGroot, Anne S.
```

```
;  TITLE OF INVENTION: HIV VACCINE CANDIDATE PEPTIDES
;  FILE REFERENCE: 17999-001
;  CURRENT APPLICATION NUMBER: US/10/200,708
;  CURRENT FILING DATE:  2002-07-22
;  PRIOR APPLICATION NUMBER: US/09/351,036
;  PRIOR FILING DATE: 1999-07-09
;  PRIOR APPLICATION NUMBER: 60/092,346
;  PRIOR FILING DATE: 1998-07-10
;  PRIOR APPLICATION NUMBER: 60/115,145
;  PRIOR FILING DATE: 1999-01-08
;  PRIOR APPLICATION NUMBER: 60/130,677
;  PRIOR FILING DATE: 1999-04-23
;  NUMBER OF SEQ ID NOS: 672
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 291
;   LENGTH: 10
;   TYPE: PRT
;   ORGANISM: Human immunodeficiency virus
US-10-200-708-291

  Query Match             32.5%;  Score 27;  DB 14;  Length 10;
  Best Local Similarity   57.1%;  Pred. No. 2.9e+02;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          6 QMTYKVY 12
              | ||::|
Db          4 QWTYQIY 10


RESULT 15
US-10-246-581-24
; Sequence 24, Application US/10246581
; Publication No. US20030097680A1
; GENERAL INFORMATION:
;  APPLICANT: Bowen, Benjamin A.
;  APPLICANT:  Chamberlin, Mark A.
;  APPLICANT:  Drummond, Bruce J.
;  APPLICANT:  McElver, John A.
;  APPLICANT:  Rothstein, Rodney J.
;  TITLE OF INVENTION: RAD51 Polypeptides
;  FILE REFERENCE: 0556D
;  CURRENT APPLICATION NUMBER: US/10/246,581
;  CURRENT FILING DATE:  2002-09-18
;  PRIOR APPLICATION NUMBER: US 09/246,963
;  PRIOR FILING DATE: 1999-02-09
;  PRIOR APPLICATION NUMBER: US 60/074743
;  PRIOR FILING DATE: 1998-02-13
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 24
;   LENGTH: 13
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: protein sequence for GFPm to ZmRAD51B fusion,
;   OTHER INFORMATION: inlcuding isoleucine and histidine linker
US-10-246-581-24

  Query Match             32.5%;  Score 27;  DB 14;  Length 13;
  Best Local Similarity   50.0%;  Pred. No. 3.9e+02;
  Matches    4;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          9 YKVYMSGT 16
              ||::|| :
Db          4 YKIHMSSS 11


Search completed: August 12, 2004, 07:10:54
Job time : 41 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:51:29 ; Search time 18 Seconds
(without alignments)
14.341 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 24558

Minimum DB seq length: 0
Maximum DB seq length: 5

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	13	61.9	5	1	US-08-193-977-16	Sequence 16, Appl
2	13	61.9	5	6	5217869-19	Patent No. 5217869
3	12	57.1	4	1	US-07-969-305-66	Sequence 66, Appl
4	12	57.1	4	1	US-07-969-305-67	Sequence 67, Appl
5	12	57.1	4	1	US-08-456-424-124	Sequence 124, App
6	12	57.1	4	2	US-08-429-964-25	Sequence 25, Appl
7	12	57.1	4	3	US-09-173-887-3	Sequence 3, Appli
8	12	57.1	4	3	US-08-842-306B-19	Sequence 19, Appl
9	12	57.1	4	3	US-08-838-973B-17	Sequence 17, Appl
10	12	57.1	4	4	US-09-294-987-4	Sequence 4, Appli
11	12	57.1	4	4	US-09-268-163-11	Sequence 11, Appl
12	12	57.1	4	4	US-09-308-175A-8	Sequence 8, Appli
13	12	57.1	4	4	US-09-601-178-2	Sequence 2, Appli
14	12	57.1	4	4	US-09-797-543-3	Sequence 3, Appli
15	12	57.1	4	4	US-09-665-362A-18	Sequence 18, Appl
16	12	57.1	4	4	US-09-665-362A-42	Sequence 42, Appl
17	12	57.1	4	5	PCT-US93-08062-25	Sequence 25, Appl
18	12	57.1	5	1	US-08-363-475-13	Sequence 13, Appl
19	12	57.1	5	1	US-07-969-305-50	Sequence 50, Appl
20	12	57.1	5	1	US-07-969-305-51	Sequence 51, Appl
21	12	57.1	5	1	US-07-969-305-52	Sequence 52, Appl
22	12	57.1	5	1	US-08-097-938-49	Sequence 49, Appl
23	12	57.1	5	1	US-08-476-000-49	Sequence 49, Appl
24	12	57.1	5	1	US-08-472-840-49	Sequence 49, Appl
25	12	57.1	5	2	US-08-441-871-138	Sequence 138, App
26	12	57.1	5	2	US-08-476-976-49	Sequence 49, Appl
27	12	57.1	5	3	US-08-474-410-49	Sequence 49, Appl
28	12	57.1	5	3	US-09-061-768A-5	Sequence 5, Appli
29	12	57.1	5	3	US-09-142-334-29	Sequence 29, Appl
30	12	57.1	5	3	US-08-842-306B-18	Sequence 18, Appl
31	12	57.1	5	3	US-08-838-973B-16	Sequence 16, Appl
32	12	57.1	5	3	US-08-486-673B-49	Sequence 49, Appl
33	12	57.1	5	4	US-08-964-747-4	Sequence 4, Appli
34	12	57.1	5	4	US-09-562-913-4	Sequence 4, Appli
35	12	57.1	5	4	US-09-764-246-5	Sequence 5, Appli
36	12	57.1	5	6	5217869-4	Patent No. 5217869
37	11	52.4	4	1	US-08-014-979-18	Sequence 18, Appl
38	11	52.4	4	1	US-08-014-979-25	Sequence 25, Appl
39	11	52.4	4	1	US-08-142-439A-8	Sequence 8, Appli
40	11	52.4	4	2	US-08-869-477-8	Sequence 8, Appli
41	11	52.4	4	2	US-08-691-997-12	Sequence 12, Appl
42	11	52.4	4	3	US-08-859-242-18	Sequence 18, Appl
43	11	52.4	4	3	US-09-171-554-10	Sequence 10, Appl
44	11	52.4	4	3	US-09-171-554-11	Sequence 11, Appl
45	11	52.4	4	4	US-08-492-411A-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-08-193-977-16
; Sequence 16, Application US/08193977
; Patent No. 5625031
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, KEVIN R.
; APPLICANT: COLEMAN, KEVIN G.
; TITLE OF INVENTION: PEPTIDE INHIBITORS OF THE P33CDK2 AND
; TITLE OF INVENTION: P34CDC2 CELL CYCLE REGULATORY KINASES AND HUMAN
; TITLE OF INVENTION: PAPILLOMAVIRUS E7 ONCOPROTEIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: REED & ROBINS
; STREET: 635 BRYANT STREET
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/193,977
; FILING DATE: 08-FEB-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-193-977-16

Query Match 61.9%; Score 13; DB 1; Length 5;
Best Local Similarity 50.0%; Pred. No. 3e+05;
Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VIAK 5
::||
Db 2 IMAK 5

```
RESULT 2
5217869-19
;Patent No. 5217869
;    APPLICANT: KAUVAR, LAWRENCE M.
;    TITLE OF INVENTION: METHOD TO PRODUCE IMMUNODIAGNOSTIC
;REAGENTS
;    NUMBER OF SEQUENCES: 121
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/255,906
;      FILING DATE: 11-OCT-1988
;SEQ ID NO:19:
;      LENGTH: 5
5217869-19

  Query Match             61.9%;  Score 13;  DB 6;  Length 5;
  Best Local Similarity   75.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIA 4
              |:||
Db          2 SMIA 5


RESULT 3
US-07-969-305-66
; Sequence 66, Application US/07969305
; Patent No. 5609872
;  GENERAL INFORMATION:
;    APPLICANT:  AHLBORG, Niklas
;    APPLICANT:  BERZINS, Klavs
;    APPLICANT:  PERLMANN, Peter
;    TITLE OF INVENTION:  NEW PEPTIDES AND THEIR USE
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  George Mason Bldg., Washington & Prince Sts.
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/969,305
;      FILING DATE:  08-APR-1993
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 9002684-0
;      FILING DATE:  17-AUG-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crane-Feury, Sharon E
;      REGISTRATION NUMBER:  36,113
;      REFERENCE/DOCKET NUMBER:  003300-286
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  66:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-07-969-305-66

  Query Match             57.1%;  Score 12;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVI 3
              |||
Db          2 SVI 4


RESULT 4
US-07-969-305-67
; Sequence 67, Application US/07969305
; Patent No. 5609872
;  GENERAL INFORMATION:
;    APPLICANT:  AHLBORG, Niklas
;    APPLICANT:  BERZINS, Klavs
;    APPLICANT:  PERLMANN, Peter
;    TITLE OF INVENTION:  NEW PEPTIDES AND THEIR USE
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  George Mason Bldg., Washington & Prince Sts.
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/969,305
;      FILING DATE:  08-APR-1993
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 9002684-0
;      FILING DATE:  17-AUG-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crane-Feury, Sharon E
;      REGISTRATION NUMBER:  36,113
;      REFERENCE/DOCKET NUMBER:  003300-286
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  67:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-07-969-305-67

  Query Match             57.1%;  Score 12;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVI 3
              |||
Db          1 SVI 3


RESULT 5
US-08-456-424-124
; Sequence 124, Application US/08456424
; Patent No. 5807979
;  GENERAL INFORMATION:
;    APPLICANT:  SATTERTHWAIT JR., ARNOLD C.
;    APPLICANT:  ARRHENIUS, THOMAS
;    APPLICANT:  CABEZAS, EDELMIRA
;    TITLE OF INVENTION:  SYNTHETIC, STABILIZED, THREE-DIMENSION
;    TITLE OF INVENTION:  POLYPEPTIDES
;    NUMBER OF SEQUENCES:  145
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
```

```
;       STREET:  755 PAGE MILL ROAD
;       CITY:  PALO ALTO
;       STATE:  CALIFORNIA
;       COUNTRY:  USA
;       ZIP:  94304-1018
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/456,424
;       FILING DATE:
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/224,059
;       FILING DATE:
;       APPLICATION NUMBER:  US 07/866,040
;       FILING DATE:  08-APR-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BOZICEVIC, KARL
;       REGISTRATION NUMBER:  28,807
;       REFERENCE/DOCKET NUMBER:  278022000120
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-813-5600
;       TELEFAX:  415-494-0792
;       TELEX:  706141
;   INFORMATION FOR SEQ ID NO:  124:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-08-456-424-124

  Query Match             57.1%;  Score 12;  DB 1;  Length 4;
  Best Local Similarity   25.0%;  Pred. No. 3e+05;
  Matches    1;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VIAK 5
              :::|
Db          1 IVSK 4


RESULT 6
US-08-429-964-25
; Sequence 25, Application US/08429964
; Patent No. 5962243
;  GENERAL INFORMATION:
;    APPLICANT:  BROWN, MICHAEL S.
;    APPLICANT:  GOLDSTEIN, JOSEPH L.
;    APPLICANT:  REISS, YUVAL
;    APPLICANT:  JAMES, GUY L.
;    TITLE OF INVENTION:  METHODS FOR THE IDENTIFICATION OF FARNESYL
;    TITLE OF INVENTION:  TRANSFERASE INHIBITORS
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ARNOLD, WHITE & DURKEE
;      STREET:  P.O. BOX 4433
;      CITY:  HOUSTON
;      STATE:  TEXAS
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS/ASCII
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/429,964
;      FILING DATE:  27-APR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/021,625
;      FILING DATE:  16-FEB-1993
;      CLASSIFICATION:  435
;      APPLICATION NUMBER:  US 07/822,011
;      FILING DATE:  ABANDONED
;      CLASSIFICATION:  435
;      APPLICATION NUMBER:  PCT/US/91/02650
;      FILING DATE:  18-APR-1991
;      CLASSIFICATION:  435
;      APPLICATION NUMBER:  US 07/615,715
;      FILING DATE:  20-NOV-1990
;      CLASSIFICATION:  435
;      APPLICATION NUMBER:  US 07/510,706
;      FILING DATE:  18-APR-1990 (ABANDONED)
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PARKER, DAVID L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UTSD:432/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (713) 789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-429-964-25

  Query Match             57.1%;  Score 12;  DB 2;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VIA 4
              |||
Db          2 VIA 4


RESULT 7
US-09-173-887-3
; Sequence 3, Application US/09173887
; Patent No. 6245884
; GENERAL INFORMATION:
;  APPLICANT: Hook, Vivian Y.H.
;  TITLE OF INVENTION: SECRETASES RELATED TO ALZHEIMER'S DEMENTIA
;  FILE REFERENCE: P-AS 3337
;  CURRENT APPLICATION NUMBER: US/09/173,887
;  CURRENT FILING DATE: 1998-10-16
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: mammalian
US-09-173-887-3

  Query Match             57.1%;  Score 12;  DB 3;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VIA 4
              |||
Db          1 VIA 3


RESULT 8
US-08-842-306B-19
; Sequence 19, Application US/08842306B
; Patent No. 6271197
;   GENERAL INFORMATION:
;        APPLICANT: Berlin, Vivian
```

```
;                    Levin, David
;                    Ohya, Yoshikazu
;                    Damagnez, Veronique
;                    Smith, Susan
;         TITLE OF INVENTION: ASSAYS AND REAGENTS FOR IDENTIFYING
;                             ANTI-FUNGAL AGENTS, AND USES RELATED THERETO
;         NUMBER OF SEQUENCES: 48
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: FOLEY, HOAG & ELIOT LLP
;              STREET: One Post Office Square
;              CITY: Boston
;              STATE: MA
;              COUNTRY: USA
;              ZIP: 02109-2170
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: WordPad
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/08/842,306B
;              FILING DATE: 23-Apr-1997
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: US 08/771,212
;              FILING DATE: 20-DEC-1996
;              APPLICATION NUMBER: US 08/631,319
;              FILING DATE: 11-APR-1996
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Vincent, Matthew P.
;              REGISTRATION NUMBER: 36,709
;              REFERENCE/DOCKET NUMBER: MIV-074.04
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 617-832-1000
;              TELEFAX: 617-832-7000
;    INFORMATION FOR SEQ ID NO: 19:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 4 amino acids
;              TYPE: amino acid
;              STRANDEDNESS: <Unknown>
;              TOPOLOGY: linear
;         MOLECULE TYPE: peptide
;         SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-08-842-306B-19

  Query Match             57.1%;  Score 12;  DB 3;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VIA 4
              |||
Db          2 VIA 4


RESULT 9
US-08-838-973B-17
; Sequence 17, Application US/08838973B
; Patent No. 6277564
;   GENERAL INFORMATION:
;         APPLICANT: Berlin, Vivian
;                    Damagnez, Veronique
;                    Smith, Susan
;         TITLE OF INVENTION: ASSAYS AND REAGENTS FOR IDENTIFYING
;                             ANTI-FUNGAL AGENTS, AND USES RELATED THERETO
;         NUMBER OF SEQUENCES: 48
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: FOLEY, HOAG & ELIOT LLP
;              STREET: One Post Office Square
;              CITY: Boston
;              STATE: MA
;              COUNTRY: USA
;              ZIP: 02109-2170
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/08/838,973B
;              FILING DATE: 23-Apr-1997
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: US 08/631,319
;              FILING DATE: 10-APR-1996
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Vincent, Matthew P.
;              REGISTRATION NUMBER: 36,709
;              REFERENCE/DOCKET NUMBER: MIV-074.05
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 617-832-1000
;              TELEFAX: 617-832-7000
;    INFORMATION FOR SEQ ID NO: 17:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 4 amino acids
;              TYPE: amino acid
;              STRANDEDNESS: <Unknown>
;              TOPOLOGY: linear
;         MOLECULE TYPE: peptide
;         SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-08-838-973B-17

  Query Match             57.1%;  Score 12;  DB 3;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VIA 4
              |||
Db          2 VIA 4


RESULT 10
US-09-294-987-4
; Sequence 4, Application US/09294987
; Patent No. 6313268
; GENERAL INFORMATION:
;  APPLICANT:  Hook, Vivian Y.H.
;  TITLE OF INVENTION:  SECRETASES RELATED TO ALZHEIMER'S DEMENTIA
;  FILE REFERENCE:  P-AS 3515
;  CURRENT APPLICATION NUMBER: US/09/294,987
;  CURRENT FILING DATE:  1999-04-20
;  PRIOR APPLICATION NUMBER:  US 09/173,887
;  PRIOR FILING DATE:  1998-10-16
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: mammalian
US-09-294-987-4

  Query Match             57.1%;  Score 12;  DB 4;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VIA 4
              |||
Db          1 VIA 3


RESULT 11
US-09-268-163-11
; Sequence 11, Application US/09268163B
; Patent No. 6353091
; GENERAL INFORMATION:
;  APPLICANT: Lipscombe, Diane
;  APPLICANT: Schorge, Stephanie
;  TITLE OF INVENTION: HUMAN N-TYPE CALCIUM CHANNEL ISOFORM AND USES THEREOF
;  FILE REFERENCE: B1055/7000
```

```
;  CURRENT APPLICATION NUMBER: US/09/268,163B
;  CURRENT FILING DATE: 1999-03-12
;  EARLIER APPLICATION NUMBER: US 60/077,901
;  EARLIER FILING DATE: 1998-03-13
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: Conus geographus
US-09-268-163-11

  Query Match             57.1%;  Score 12;  DB 4;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VIA 4
              |||
Db          2 VIA 4


RESULT 12
US-09-308-175A-8
; Sequence 8, Application US/09308175A
; Patent No. 6355617
; GENERAL INFORMATION:
;  APPLICANT:  LUKE, Richard William Arthur
;  APPLICANT:  COTTON, Ronald
;  TITLE OF INVENTION:  PEPTIDE DERIVATIVES
;  FILE REFERENCE:  1991-174
;  CURRENT APPLICATION NUMBER: US/09/308,175A
;  CURRENT FILING DATE:  1999-05-17
;  PRIOR APPLICATION NUMBER:  PCT/GB97/03199
;  PRIOR FILING DATE:  1997-11-21
;  PRIOR APPLICATION NUMBER:  GB 9624562.6
;  PRIOR FILING DATE:  1996-11-27
;  NUMBER OF SEQ ID NOS:  22
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  synthetic
;   OTHER INFORMATION: peptide
US-09-308-175A-8

  Query Match             57.1%;  Score 12;  DB 4;  Length 4;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 IAK 5
              :||
Db          1 VAK 3


RESULT 13
US-09-601-178-2
; Sequence 2, Application US/09601178
; Patent No. 6548306
; GENERAL INFORMATION:
;  APPLICANT: ADMON, Arie
;  APPLICANT:  PALTIELI, Yoav
;  APPLICANT:  MANDELI, Silvia
;  APPLICANT:  SLOTKY, Ronit
;  TITLE OF INVENTION: PLACENTAL PROTEIN 13
;  FILE REFERENCE: ADMON=1
;  CURRENT APPLICATION NUMBER: US/09/601,178
;  CURRENT FILING DATE:  2001-01-31
;  PRIOR APPLICATION NUMBER: PCT/IL99/00036
;  PRIOR FILING DATE: 1999-01-21
;  PRIOR APPLICATION NUMBER: IL 123098
;  PRIOR FILING DATE: 1998-01-29
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-601-178-2

  Query Match             57.1%;  Score 12;  DB 4;  Length 4;
  Best Local Similarity   75.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 VIAK 5
              || |
Db          1 VIIK 4


RESULT 14
US-09-797-543-3
; Sequence 3, Application US/09797543
; Patent No. 6627409
; GENERAL INFORMATION:
;  APPLICANT: Hook, Vivian Y.H.
;  TITLE OF INVENTION: SECRETASES RELATED TO ALZHEIMER'S DEMENTIA
;  FILE REFERENCE: P-AS 4579
;  CURRENT APPLICATION NUMBER: US/09/797,543
;  CURRENT FILING DATE:  2001-05-29
;  PRIOR APPLICATION NUMBER: US 09/173,887
;  PRIOR FILING DATE: 1998-10-16
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: mammalian
US-09-797-543-3

  Query Match             57.1%;  Score 12;  DB 4;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VIA 4
              |||
Db          1 VIA 3


RESULT 15
US-09-665-362A-18
; Sequence 18, Application US/09665362A
; Patent No. 6632626
; GENERAL INFORMATION:
;  APPLICANT: BROWN, MICHAEL S.
;  APPLICANT:  GOLDSTEIN, JOSEPH L.
;  APPLICANT:  REISS, YUVAL
;  TITLE OF INVENTION: METHODS  OF ASSAYING FARNESYL TRANSFERASE
;  FILE REFERENCE: UTSD:249USD1
;  CURRENT APPLICATION NUMBER: US/09/665,362A
;  CURRENT FILING DATE:  2003-07-22
;  PRIOR APPLICATION NUMBER: 07/937,893
;  PRIOR FILING DATE: 1992-12-22
;  PRIOR APPLICATION NUMBER: 07/615,715
;  PRIOR FILING DATE: 1990-11-20
;  PRIOR APPLICATION NUMBER: 07/510,706
;  PRIOR FILING DATE: 1990-04-18
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
;   LENGTH: 4
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Synthetic
;   OTHER INFORMATION: Peptide
```

US-09-665-362A-18

Query Match 57.1%; Score 12; DB 4; Length 4;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          2 VIA 4
              |||
Db          2 VIA 4
```

Search completed: August 12, 2004, 06:55:49
Job time : 19 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 1.56762 Seconds
(without alignments)
1043.144 Million cell updates/sec

Title: US-09-890-463-2
Perfect score: 83
Sequence: 1 SVIAKQMTYKVYMSGTV 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	54.2	398	2	G96956	protein of short-c
2	43	51.8	424	2	T43498	hypothetical prote
3	43	51.8	3770	2	A40889	delta-(L-alpha-ami
4	41	49.4	3587	2	I40486	surfactin syntheta
5	41	49.4	3588	2	I40485	surfactin syntheta
6	40	48.2	218	2	E86687	hypothetical prote
7	40	48.2	263	2	C97225	hypothetical prote
8	40	48.2	285	2	F83890	sugar transport sy
9	40	48.2	294	2	G83962	hypothetical prote
10	40	48.2	870	2	T10634	hypothetical prote
11	40	48.2	1086	2	AH2136	microcystin synthe
12	39	47.0	161	2	A40745	odorant receptor (
13	39	47.0	170	2	S64488	regulatory protein
14	39	47.0	277	2	A70233	hypothetical prote
15	39	47.0	337	2	G70394	PlsX protein - Aqu
16	39	47.0	346	2	H71473	probable leucine d
17	39	47.0	368	2	F81816	phosphoserine tran
18	39	47.0	368	2	H81059	phosphoserine amin
19	39	47.0	459	2	T29945	hypothetical prote
20	39	47.0	786	2	AG0182	probable membrane
21	39	47.0	846	2	S59262	hypothetical prote
22	39	47.0	3712	1	YGCEVC	alpha-aminoadipyl-
23	38	45.8	140	2	A69445	hypothetical prote
24	38	45.8	192	2	AI0383	probable lipoprote
25	38	45.8	192	2	AH0557	probable lipoprote
26	38	45.8	226	2	D85540	probable polymeras
27	38	45.8	226	2	B64773	lipoprotein yajG p
28	38	45.8	226	2	H90689	probable polymeras
29	38	45.8	227	2	E25973	pertussis toxin ch
30	38	45.8	234	2	F95952	probable membrane-
31	38	45.8	304	2	B36716	protochlorophyllid
32	38	45.8	353	1	KIBEMV	thymidine kinase (
33	38	45.8	554	2	H70011	exo-alpha-1,4-gluc
34	38	45.8	560	2	S38035	probable serine/th
35	38	45.8	635	2	S73358	topoisomerase IV c
36	37	44.6	227	1	F25973	pertussis toxin ch
37	37	44.6	351	2	E86605	Leucine dehydrogen
38	37	44.6	351	2	F72020	leucine dehydrogen
39	37	44.6	384	2	G81436	probable serine/th
40	37	44.6	440	2	F84955	UDP-N-acetylmuramo
41	37	44.6	478	1	S03826	UMP synthase - sli
42	37	44.6	552	2	B86899	hypothetical prote
43	37	44.6	563	2	S10176	protochlorophyllid
44	37	44.6	1184	2	T41515	coiled coil protei
45	37	44.6	2560	1	I40457	peptide synthetase

ALIGNMENTS

RESULT 1
G96956
protein of short-chain alcohol dehydrogenase family [imported] - Clostridium acetobutylic
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C;Accession: G96956
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clos
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: G96956
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-398 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK78442.1; PID:g15023320; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0462
C;Superfamily: Xylella fastidiosa hypothetical protein XF1835

Query Match 54.2%; Score 45; DB 2; Length 398;
Best Local Similarity 47.1%; Pred. No. 3.2;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

```
Qy          1 SVIAKQMTYKVYMSGTV 17
              | |    |||:|  ||:
Db        224 SYIGSPRTYKIYREGTI 240
```

RESULT 2
T43498
hypothetical protein DKFZp586C1021.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C;Accession: T43498
R;Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, December 1999
A;Reference number: Z22515
A;Accession: T43498
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-424 <AAA>
A;Cross-references: EMBL:AL133640
A;Experimental source: adult uterus; clone DKFZp586C1021
C;Genetics:
A;Note: DKFZp586C1021.1

Query Match 51.8%; Score 43; DB 2; Length 424;
Best Local Similarity 43.8%; Pred. No. 7.8;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 SVIAKQMTYKVYMSGT 16
              |::|   :|:|| :||
Db          1 SIVALNKSYEVYFTGT 16
```

RESULT 3
A40889
delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase - Emericella nidulans
N;Alternate names: ACV synthetase
C;Species: Emericella nidulans, Aspergillus nidulans
C;Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 03-Nov-2000
C;Accession: A40889; S16466
R;MacCabe, A.P.; van Liempt, H.; Palissa, H.; Unkles, S.E.; Riach, M.B.R.; Pfeifer, E.;
J. Biol. Chem. 266, 12646-12654, 1991
A;Title: delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase from Aspergillus ni
thway.
A;Reference number: A40889; MUID:91286299; PMID:2061333
A;Accession: A40889
A;Molecule type: DNA
A;Residues: 1-3770 <MAC>
A;Cross-references: GB:X54853; NID:g2318; PIDN:CAA38631.1; PID:g2319
A;Note: the sequence of residues 3129-3148 and the corresponding nucleotide sequence are
C;Genetics:
A;Gene: acvA
C;Superfamily: alpha-aminoadipyl-cysteinyl-valine synthetase; acetate-CoA ligase homolog
C;Keywords: antibiotic biosynthesis; carrier protein; penicillin biosynthesis; phosphopa
F;365-831/Domain: acetate-CoA ligase homology <ACL1>
F;848-918/Domain: acyl carrier protein homology <ACP1>
F;1458-1915/Domain: acetate-CoA ligase homology <ACL2>
F;1931-2001/Domain: acyl carrier protein homology <ACP2>
F;2539-3001/Domain: acetate-CoA ligase homology <ACL3>
F;3018-3086/Domain: acyl carrier protein homology <ACP3>
F;882,1965,3050/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match   51.8%; Score 43; DB 2; Length 3770;
  Best Local Similarity   50.0%; Pred. No. 71;
  Matches   8; Conservative   3; Mismatches   5; Indels   0; Gaps   0;

```
Qy          1 SVIAKQMTYKVYMSGT 16
              |: :||: |  | |||
Db        468 SLTSKQLAYVTYTSGT 483
```

RESULT 4
I40486
surfactin synthetase component II - Bacillus subtilis
N;Alternate names: surfactin synthetase srfA2; surfactin synthetase/competence protein s
N;Contains: acid-amino-acid ligase (EC 6.3.2.-)
C;Species: Bacillus subtilis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 03-Nov-2000
C;Accession: I40486; S60866; C69718; S46968; S35518; S25658; S34986
R;Cosmina, P.; Rodriguez, F.; de Ferra, F.; Grandi, G.; Perego, M.; Venema, G.; van Sind
Mol. Microbiol. 8, 821-831, 1993
A;Title: Sequence and analysis of the genetic locus responsible for surfactin synthesis
A;Reference number: I40485; MUID:93360813; PMID:8355609
A;Accession: I40486
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3587 <RES>
A;Cross-references: EMBL:X70356; NID:g396480; PIDN:CAA49817.1; PID:g396482
A;Experimental source: strain W168 derivative of JH642
R;Hamoen, L.W.; Eshuis, H.; Jongbloed, J.; Venema, G.; van Sinderen, D.
Mol. Microbiol. 15, 55-63, 1995
A;Title: A small gene, designated comS, located within the coding region of the fourth a
A;Reference number: S60866; MUID:95272393; PMID:7752896
A;Accession: S60866
A;Molecule type: DNA
A;Residues: 977-1104 <HAM>
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: C69718
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-3587 <KUN>
A;Cross-references: GB:Z99105; GB:AL009126; NID:g2632457; PIDN:CAB12143.1; PID:g2632635
A;Experimental source: strain 168
R;Fabret, C.; Quentin, Y.; Guiseppi, A.; Busuttil, J.; Haiech, J.; Denizot, F.
submitted to the EMBL Data Library, March 1993
A;Reference number: S46967
A;Accession: S46968
A;Molecule type: DNA
A;Residues: 1-32,'F',34-41,'G',43-109,'D',111-114,'G',116-138,'V',140-258,'W',260-308,'A
1756-1914,'PK',1917-2138,'SRL',2142,'DSLN',2146-2444,'Q',2446-2712,'H',2714-2722,'H',272
A;Cross-references: EMBL:X72672; NID:g516358; PIDN:CAA51223.1; PID:g516360
R;Fuma, S.; Fujishima, Y.; Corbell, N.; D'Souza, C.; Nakano, M.M.; Zuber, P.; Yamane, K.
Nucleic Acids Res. 21, 93-97, 1993
A;Title: Nucleotide sequence of 5' portion of srfA that contains the region required for
A;Reference number: S35517; MUID:93181186; PMID:8441623
A;Accession: S35518
A;Status: significant sequence differences
A;Molecule type: DNA
A;Cross-references: EMBL:D13262; NID:g216345; PID:g216347
A;Experimental source: strain 168 trpC2
R;Borchert, S.; Patil, S.S.; Marahiel, M.A.
FEMS Microbiol. Lett. 92, 175-180, 1992
A;Title: Identification of putative multifunctional peptide synthetase genes using highl
A;Reference number: S25658
A;Accession: S25658
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 168,'C',170-171,281-283;514-595,597-647,'R',649-679,'ETL',683-693,'DKR',697,
A;Cross-references: EMBL:X65835; NID:g40202; PIDN:CAA46678.1; PID:g40203
A;Experimental source: strain ATCC 21332
C;Comment: This protein contains several amino acid-activating domains for the synthesis
the amino-terminal region of this protein, appear to be required for the development of
C;Genetics:
A;Gene: srfAB; srfA2
C;Superfamily: surfactin synthetase; acetate-CoA ligase homology; acyl carrier protein h
C;Keywords: antibiotic biosynthesis; carrier protein; duplication; ligase; phosphopanteth
F;511-951/Domain: acetate-CoA ligase homology <ACL1>
F;968-1035/Domain: acyl carrier protein homology <ACP1>
F;1036-1481/Domain: repeat <RPT1>
F;1542-1995/Domain: acetate-CoA ligase homology <ACL2>
F;2013-2081/Domain: acyl carrier protein homology <ACP2>
F;2082-2529/Domain: repeat <RPT2>
F;2591-3024/Domain: acetate-CoA ligase homology <ACL3>
F;3041-3108/Domain: acyl carrier protein homology <ACP3>
F;999,2045,3073/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match   49.4%; Score 41; DB 2; Length 3587;
  Best Local Similarity   43.8%; Pred. No. 1.6e+02;
  Matches   7; Conservative   4; Mismatches   5; Indels   0; Gaps   0;

```
Qy          1 SVIAKQMTYKVYMSGT 16
              :| |: : | :| |||
Db       2687 AVTAENLAYMIYTSGT 2702
```

RESULT 5
I40485
surfactin synthetase component I - Bacillus subtilis
N;Alternate names: competence protein srfAA; surfactin production protein srfAA; surfact

C;Species: Bacillus subtilis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 03-Nov-2000
C;Accession: I40485; B69718; S35517; A37323; S46967; A43705; S34985
R;Cosmina, P.; Rodriguez, F.; de Ferra, F.; Grandi, G.; Perego, M.; Venema, G.; van Sind
Mol. Microbiol. 8, 821-831, 1993
A;Title: Sequence and analysis of the genetic locus responsible for surfactin synthesis
A;Reference number: I40485; MUID:93360813; PMID:8355609
A;Accession: I40485
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3588 <RES>
A;Cross-references: EMBL:X70356; NID:g396480; PIDN:CAA49816.1; PID:g396481
A;Experimental source: strain W168 derivative of JH642
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: B69718
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-3588 <KUN>
A;Cross-references: GB:Z99105; GB:AL009126; NID:g2632457; PIDN:CAB12142.1; PID:g2632634
A;Experimental source: strain 168
R;Fuma, S.; Fujishima, Y.; Corbell, N.; D'Souza, C.; Nakano, M.M.; Zuber, P.; Yamane, K.
Nucleic Acids Res. 21, 93-97, 1993
A;Title: Nucleotide sequence of 5' portion of srfA that contains the region required for
A;Reference number: S35517; MUID:93181186; PMID:8441623
A;Accession: S35517
A;Status: nucleic acid sequence not shown; significant sequence differences
A;Molecule type: DNA
A;Cross-references: EMBL:D13262; NID:g216345; PID:g216346
A;Experimental source: strain 168 trpC2
A;Note: protein sequence not complete, the nucleotide sequence was submitted to the EMBL
R;Nakano, M.M.; Magnuson, R.; Myers, A.; Curry, J.; Grossman, A.D.; Zuber, P.
J. Bacteriol. 173, 1770-1778, 1991
A;Title: srfA is an operon required for surfactin production, competence development, an
A;Reference number: A37323; MUID:91154134; PMID:1847909
A;Accession: A37323
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-46,'EV',49-145,'I',147-150,'L',152,'AN',155-280,'T',282-307,'SF',310-384 <
A;Cross-references: GB:M59939; NID:g14366
R;Nakano, M.M.; Xia, L.; Zuber, P.
J. Bacteriol. 173, 5487-5493, 1991
A;Title: Transcription initiation region of the srfA operon, which is controlled by the
A;Reference number: A43705; MUID:91358326; PMID:1715856
A;Contents: annotation
R;Fabret, C.; Quentin, Y.; Guiseppi, A.; Busuttil, J.; Haiech, J.; Denizot, F.
submitted to the EMBL Data Library, March 1993
A;Reference number: S46967
A;Accession: S46967
A;Molecule type: DNA
A;Residues: 3249-3271,'A',3273-3316,'R',3318-3451,'Y',3453-3483,'DE',3486-3487,'DAGL',34
A;Cross-references: EMBL:X72672; NID:g516358; PIDN:CAA51222.1; PID:g516359
A;Experimental source: strain 168 trpC2
C;Comment: This enzyme is one of several in the multienzyme complex that synthesizes the
nce.
C;Genetics:
A;Gene: srfAA; srfA1
A;Note: srfAA is the first gene of the srfA operon
C;Superfamily: surfactin synthetase; acetate-CoA ligase homology; acyl carrier protein h
C;Keywords: antibiotic biosynthesis; carrier protein; phosphopantetheine; phosphoprotein
F;507-956/Domain: acetate-CoA ligase homology <ACL1>
F;974-1042/Domain: acyl carrier protein homology <ACP1>
F;1043-1488/Domain: repeat <RPT1>
F;1549-1993/Domain: acetate-CoA ligase homology <ACL2>
F;2011-2079/Domain: acyl carrier protein homology <ACP2>
F;2080-2527/Domain: repeat <RPT2>
F;2589-3025/Domain: acetate-CoA ligase homology <ACL3>
F;3042-3109/Domain: acyl carrier protein homology <ACP3>
F;1006,2043,3074/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match 49.4%; Score 41; DB 2; Length 3588;
  Best Local Similarity 43.8%; Pred. No. 1.6e+02;
  Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGT 16
:| |: : | :| |||
Db 2688 AVTAENLAYMIYTSGT 2703

RESULT 6
E86687
hypothetical protein [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: E86687
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlic
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: E86687
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-218 <STO>
A;Cross-references: GB:AE005176; PID:g12723383; PIDN:AAK04599.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: yfaA

  Query Match 48.2%; Score 40; DB 2; Length 218;
  Best Local Similarity 70.0%; Pred. No. 14;
  Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 5 KQMTYKVYMS 14
|:|||| |:|
Db 100 KEMTYKFYIS 109

RESULT 7
C97225
hypothetical protein CAC2643 [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: C97225
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clos
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: C97225
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-263 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK80590.1; PID:g15025670; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2643

  Query Match 48.2%; Score 40; DB 2; Length 263;
  Best Local Similarity 43.8%; Pred. No. 17;
  Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGT 16

```
                | : : :||| | ||:
Db        152 SGVVQDITYKAYTSGS 167


RESULT 8
F83890
sugar transport system (permease) BH1926 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: F83890
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: F83890
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-285 <STO>
A;Cross-references: GB:AP001513; GB:BA000004; NID:g10174345; PIDN:BAB05645.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH1926
C;Superfamily: maltose transport protein malG

  Query Match             48.2%;  Score 40;  DB 2;  Length 285;
  Best Local Similarity   61.5%;  Pred. No. 18;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          4 AKQMTYKVYMSGT 16
              |:|| | : ||||
Db        244 ARQMDYGMIMSGT 256


RESULT 9
G83962
hypothetical protein BH2503 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: G83962
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: G83962
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-294 <STO>
A;Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BAB06222.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH2503
C;Superfamily: conserved hypothetical protein HI1714

  Query Match             48.2%;  Score 40;  DB 2;  Length 294;
  Best Local Similarity   46.2%;  Pred. No. 19;
  Matches    6;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          5 KQMTYKVYMSGTV 17
              :|: ||||:: |:
Db        133 EQLGYKVYLTSTI 145


RESULT 10
T10634
hypothetical protein T13K14.80 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 18-Aug-2000
C;Accession: T10634
R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16991
A;Accession: T10634
A;Molecule type: DNA
A;Residues: 1-870 <BEV>
A;Cross-references: EMBL:AL080282; GSPDB:GN00062; ATSP:T13K14.80
A;Experimental source: cultivar Columbia; BAC clone T13K14
C;Genetics:
A;Gene: ATSP:T13K14.80
A;Map position: 4
A;Introns: 72/3; 105/3; 257/2; 328/2; 635/1; 685/3; 724/3
C;Superfamily: Arabidopsis thaliana hypothetical protein T13K14.70

  Query Match             48.2%;  Score 40;  DB 2;  Length 870;
  Best Local Similarity   53.3%;  Pred. No. 56;
  Matches    8;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSG 15
              | :||:  | || ||
Db        803 SKLAKRHNYSVYFSG 817


RESULT 11
AH2136
microcystin synthetase B [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AH2136
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AH2136
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1086 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB74346.1; PID:g17131740; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all2647
C;Superfamily: peptide synthetase ppsE; acetate-CoA ligase homology; acyl carrier protei
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;1015/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             48.2%;  Score 40;  DB 2;  Length 1086;
  Best Local Similarity   46.7%;  Pred. No. 71;
  Matches    7;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 VIAKQMTYKVYMSGT 16
              |  :|:|| :| ||:
Db        610 VTPEQLTYLIYTSGS 624


RESULT 12
A40745
odorant receptor (clone M50) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 21-Sep-1994 #sequence_revision 18-Nov-1994 #text_change 26-Aug-1999
C;Accession: A40745
R;Ressler, K.J.; Sullivan, S.L.; Buck, L.B.
Cell 73, 597-609, 1993
A;Title: A zonal organization of odorant receptor gene expression in the olfactory epithe
A;Reference number: A40745; MUID:93258822; PMID:7683976
A;Accession: A40745
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-161 <RES>
A;Cross-references: GB:L14568; NID:g293757; PIDN:AAA39852.1; PID:g293758
A;Experimental source: olfactory epithelium
A;Note: sequence extracted from NCBI backbone (NCBIP:131747)
C;Superfamily: olfactory receptor OR14
C;Keywords: G protein-coupled receptor; transmembrane protein

  Query Match             47.0%;  Score 39;  DB 2;  Length 161;
  Best Local Similarity   66.7%;  Pred. No. 16;
```

```
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 IAKQMTYKVYMS 14
              | | :|||| ||
Db          5 ICKPLTYKVIMS 16


RESULT 13
S64488
regulatory protein CBP4 precursor - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein G7122; protein YGR174c
C;Species: Saccharomyces cerevisiae
C;Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 21-Jul-2000
C;Accession: S64488; A53928
R;Hebling, U.; Hofmann, B.; Delius, H.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64003
A;Accession: S64488
A;Molecule type: DNA
A;Residues: 1-170 <HEB>
A;Cross-references: EMBL:Z72959; NID:g1323307; PID:e243557; PID:g1323308; MIPS:YGR174c
A;Experimental source: strain S288C
R;Crivellone, M.D.
J. Biol. Chem. 269, 21284-21292, 1994
A;Title: Characterization of CBP4, a new gene essential for the expression of ubiquinol-
A;Reference number: A53928; MUID:94342301; PMID:8063753
A;Accession: A53928
A;Molecule type: DNA
A;Residues: 1-64,'F',66-170 <CRI>
A;Cross-references: GB:U10700; NID:g505645; PIDN:AAA61566.1; PID:g505646
C;Genetics:
A;Gene: SGD:CBP4
A;Cross-references: SGD:S0003406; MIPS:YGR174c
A;Map position: 7R
A;Genome: nuclear
C;Keywords: mitochondrion; transmembrane protein
F;30-51/Domain: transmembrane #status predicted <TMM>

  Query Match             47.0%;  Score 39;  DB 2;  Length 170;
  Best Local Similarity   61.5%;  Pred. No. 17;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYM 13
              :|||||  || |:
Db         10 AVIAKQRQYKHYL 22


RESULT 14
A70233
hypothetical protein BBG17 - Lyme disease spirochete plasmid G/lp28-2
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 28-Jul-2000
C;Accession: A70233
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: A70233
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-277 <KLE>
A;Cross-references: GB:AE000786; NID:g2690008; PIDN:AAC66065.1; PID:g2690022; TIGR:BBG17
A;Experimental source: strain B31
C;Genetics:
A;Genome: plasmid
C;Superfamily: Borrelia burgdorferi hypothetical protein BBG17

  Query Match             47.0%;  Score 39;  DB 2;  Length 277;
  Best Local Similarity   41.2%;  Pred. No. 27;
  Matches    7;  Conservative    5;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |:  |:| ||:::  ||
Db         70 SIFFKEMAYKMHVFDTV 86


RESULT 15
G70394
PlsX protein - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 28-Jul-2003
C;Accession: G70394
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: G70394
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-337 <AQF>
A;Cross-references: GB:AE000723; NID:g2983569; PIDN:AAC07145.1; PID:g2983573; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: plsX
C;Superfamily: phospholipid biosynthesis protein, PlsX type

  Query Match             47.0%;  Score 39;  DB 2;  Length 337;
  Best Local Similarity   50.0%;  Pred. No. 33;
  Matches    6;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 AKQMTYKVYMSG 15
              ||:: ||:|: |
Db         26 AKELGYKIYLVG 37


Search completed: August 12, 2004, 06:13:50
Job time : 3.56762 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 1.04508 Seconds
(without alignments)
847.008 Million cell updates/sec

Title: US-09-890-463-2
Perfect score: 83
Sequence: 1 SVIAKQMTYKVYMSGTV 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	43	51.8	3770	1	ACVS_EMENI	P27742	emericella
2	41	49.4	3587	1	SRF2_BACSU	Q04747	bacillus su
3	41	49.4	3588	1	SRF1_BACSU	P27206	bacillus su
4	40	48.2	294	1	ENGC_BACHD	Q9k9z1	bacillus ha
5	40	48.2	697	1	SM2A_SCHGR	Q9xzc8	schistocerc
6	39	47.0	161	1	OL7A_MOUSE	P34985	mus musculu
7	39	47.0	170	1	CBP4_YEAST	P37267	saccharomyc
8	39	47.0	337	1	PLSX_AQUAE	O67186	aquifex aeo
9	39	47.0	368	1	SERC_NEIMA	O34370	neisseria m
10	39	47.0	368	1	SERC_NEIMB	P57007	neisseria m
11	39	47.0	846	1	SP98_YEAST	P53540	saccharomyc
12	39	47.0	3712	1	ACVS_CEPAC	P25464	cephalospor
13	38	45.8	140	1	YF62_ARCFU	O28710	archaeoglob
14	38	45.8	192	1	YAJG_ECOLI	P36671	escherichia
15	38	45.8	304	1	BCHL_RHOCA	P26237	rhodobacter
16	38	45.8	352	1	KITH_HSVMD	P17653	marek's dis
17	38	45.8	386	1	MTLD_OCEIH	Q8en87	oceanobacil
18	38	45.8	560	1	PTK1_YEAST	P36002	saccharomyc
19	38	45.8	635	1	PARE_MYCPN	P78016	mycoplasma
20	37	44.6	440	1	MURD_BUCAI	P57313	buchnera ap
21	37	44.6	444	1	CHLB_CHLPT	P37824	chlamydomon
22	37	44.6	478	1	PYR5_DICDI	P09556	dictyosteli
23	37	44.6	563	1	CHLB_CHLMO	P17652	chlamydomon
24	37	44.6	2560	1	PPS2_BACSU	P39846	bacillus su
25	36	43.4	134	1	YA7B_HAEIN	Q57425	haemophilus
26	36	43.4	155	1	H2B4_VOLCA	P16868	volvox cart
27	36	43.4	157	1	H2B3_VOLCA	P16867	volvox cart
28	36	43.4	172	1	FABA_VIBPA	Q87pc5	vibrio para
29	36	43.4	172	1	FABA_VIBVU	Q8d9h3	vibrio vuln
30	36	43.4	190	1	KCY_THEVO	Q97bv0	thermoplasm
31	36	43.4	232	1	RHO6_HUMAN	Q92730	homo sapien
32	36	43.4	297	1	BCHL_RHOSH	Q9rfd6	rhodobacter
33	36	43.4	349	1	IL8A_RAT	P70612	rattus norv
34	36	43.4	379	1	PURK_BACSU	P12045	bacillus su
35	36	43.4	384	1	YH74_VIBCH	Q9kr69	vibrio chol
36	36	43.4	453	1	MPL_HAEIN	P43948	haemophilus
37	36	43.4	489	1	SYE_OCEIH	Q8eu02	oceanobacil
38	36	43.4	517	1	MBI3_YEAST	Q9zzw7	saccharomyc
39	36	43.4	1224	1	RPOD_PINTH	P41606	pinus thunb
40	36	43.4	6548	1	EPPL_MOUSE	Q8r0w0	mus musculu
41	35	42.2	60	1	CX3_NAJHA	P01459	naja haje a
42	35	42.2	60	1	CX4_NAJHA	P01461	naja haje a
43	35	42.2	60	1	CX8_NAJHA	P01460	naja haje a
44	35	42.2	100	1	SUI1_THEVO	Q97bw9	thermoplasm
45	35	42.2	171	1	FABA_PSEAE	O33877	pseudomonas

ALIGNMENTS

RESULT 1
ACVS_EMENI
ID ACVS_EMENI STANDARD; PRT; 3770 AA.
AC P27742;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE N-(5-amino-5-carboxypentanoyl)-L-cysteinyl-D-valine synthase
DE (EC 6.3.2.26) (Delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine
DE synthetase) (ACV synthetase) (ACVS).
GN ACVA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=G191;
RX MEDLINE=91286299; PubMed=2061333;
RA Maccabe A.P., van Liempt H., Pallissa H., Unkles S.E., Riach M.B.R.,
RA Pfeifer E., von Doehren H., Kinghorn J.R.;
RT "Delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase from
RT Aspergillus nidulans. Molecular characterization of the acvA gene
RT encoding the first enzyme of the penicillin biosynthetic pathway.";
RL J. Biol. Chem. 266:12646-12654(1991).
CC -!- FUNCTION: Each of the constituent amino acids of the tripeptide
CC acv are activated as aminoacyl-adenylates with peptide bonds
CC formed through the participation of amino acid thiolester
CC intermediates.
CC -!- CATALYTIC ACTIVITY: L-2-aminohexanedioate + L-cysteine + L-valine
CC + 3 ATP = N-[L-5-amino-5-carboxypentanoyl]-L-cysteinyl-D-valine +
CC 3 AMP + 3 diphosphate.
CC -!- COFACTOR: Contains 3 covalently bound phosphopantetheines
CC (Potential).
CC -!- PATHWAY: Biosynthesis of penicillin and cephalosporin; first step.
CC -!- PTM: The N-terminus is blocked.
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
CC -!- SIMILARITY: Contains 3 acyl carrier domains.

DR EMBL; X54853; CAA38631.1; -.
DR PIR; A40889; A40889.
DR HSSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR006163; Pp_bind.
DR InterPro; IPR006162; Ppantne_S.
DR InterPro; IPR000379; Ser_estrs.

```
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00501; AMP-binding; 3.
DR   Pfam; PF00668; Condensation; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00455; AMP_BINDING; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Ligase; Antibiotic biosynthesis; Multifunctional enzyme;
KW   Repeat; Phosphopantetheine.
FT   REPEAT      321    910       DOMAIN 1 (ADIPATE-ACTIVATING).
FT   REPEAT     1413   1993       DOMAIN 2 (CYSTEINE-ACTIVATING).
FT   REPEAT     2494   3078       DOMAIN 3 (VALINE-ACTIVATING).
FT   DOMAIN      850    919       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1929   2002       ACYL CARRIER (ACP) 2.
FT   DOMAIN     3020   3087       ACYL CARRIER (ACP) 3.
FT   BINDING     882    882       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1965   1965       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    3050   3050       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   3623   3623       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   3770 AA;  422448 MW;  CB66B6D232A58CB0 CRC64;

  Query Match             51.8%;  Score 43;  DB 1;  Length 3770;
  Best Local Similarity   50.0%;  Pred. No. 28;
  Matches    8;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              |: :||: |  | |||
Db        468 SLTSKQLAYVTYTSGT 483


RESULT 2
SRF2_BACSU
ID   SRF2_BACSU     STANDARD;      PRT;  3587 AA.
AC   Q04747;
DT   01-FEB-1995 (Rel. 31, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Surfactin synthetase subunit 2.
GN   SRFAB OR SRFA2 OR COML OR BSU03490.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE OF 1-3077 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=93181186; PubMed=8441623;
RA   Fuma S., Fujishima Y., Corbell N., D'Souza C., Nakano M.M.,
RA   Zuber P., Yamane K.;
RT   "Nucleotide sequence of 5' portion of srfA that contains the region
RT   required for competence establishment in Bacillus subtilis.";
RL   Nucleic Acids Res. 21:93-97(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / JH642;
RX   MEDLINE=93360813; PubMed=8355609;
RA   Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,
RA   Venema G., van Sinderen D.;
RT   "Sequence and analysis of the genetic locus responsible for surfactin
RT   synthesis in Bacillus subtilis.";
RL   Mol. Microbiol. 8:821-831(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=97124189; PubMed=8969502;
RA   Yamane K., Kumano M., Kurita K.;
RT   "The 25 degrees-36 degrees region of the Bacillus subtilis
RT   chromosome: determination of the sequence of a 146 kb segment and
RT   identification of 113 genes.";
RL   Microbiology 142:3047-3056(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [5]
RP   SEQUENCE OF 514-800 FROM N.A.
RC   STRAIN=ATCC 21332;
RX   MEDLINE=92290255; PubMed=1601288;
RA   Borchert S., Patil S.S., Marahiel M.A.;
RT   "Identification of putative multifunctional peptide synthetase genes
RT   using highly conserved oligonucleotide sequences derived from known
RT   synthetases.";
RL   FEMS Microbiol. Lett. 71:175-180(1992).
CC   -!- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
CC       ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
CC       ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
CC   -!- COFACTOR: Contains 3 covalently bound phosphopantetheines.
CC   -!- PATHWAY: Cyclic peptide antibiotic surfactin biosynthesis.
CC   -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC       family.
CC   -!- SIMILARITY: Contains 3 acyl carrier domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13262; BAA02523.1; -.
DR   EMBL; X70356; CAA49817.1; -.
DR   EMBL; D50453; BAA08983.1; -.
DR   EMBL; Z99105; CAB12143.1; -.
DR   EMBL; X65835; CAA46678.1; -.
DR   PIR; I40486; I40486.
DR   HSSP; P14687; 1AMU.
DR   SubtiList; BG10169; srfAB.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_S.
DR   Pfam; PF00501; AMP-binding; 3.
DR   Pfam; PF00668; Condensation; 4.
```

```
DR   Pfam; PF00550; pp-binding; 3.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00455; AMP_BINDING; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Ligase; Antibiotic biosynthesis; Phosphopantetheine; Sporulation;
KW   Multifunctional enzyme; Repeat; Complete proteome.
FT   REPEAT        ?   1040       DOMAIN 1 (VAL-ACTIVATING).
FT   REPEAT        ?   2096       DOMAIN 2 (ASP-ACTIVATING).
FT   REPEAT        ?   3114       DOMAIN 3 (D-LEU-ACTIVATING).
FT   DOMAIN      970   1036       ACYL CARRIER (ACP) 1.
FT   DOMAIN     2015   2082       ACYL CARRIER (ACP) 2.
FT   DOMAIN     3043   3109       ACYL CARRIER (ACP) 3.
FT   BINDING     999    999       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2045   2045       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    3073   3073       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT     33     33       S -> F (IN REF. 1).
FT   CONFLICT     42     42       A -> G (IN REF. 1).
FT   CONFLICT    110    110       Q -> D (IN REF. 1).
FT   CONFLICT    113    115       RQA -> AQG (IN REF. 1).
FT   CONFLICT    139    139       A -> V (IN REF. 1).
FT   CONFLICT    259    259       L -> W (IN REF. 1).
FT   CONFLICT    309    309       R -> A (IN REF. 1).
FT   CONFLICT    478    480       TPA -> SRP (IN REF. 1).
FT   CONFLICT    596    596       MISSING (IN REF. 5).
FT   CONFLICT    648    648       A -> R (IN REF. 1).
FT   CONFLICT    680    682       RHV -> ETL (IN REF. 1).
FT   CONFLICT    694    698       EQSIT -> DKRIS (IN REF. 5).
FT   CONFLICT    788    788       M -> L (IN REF. 5).
FT   CONFLICT    939    940       PL -> LV (IN REF. 1).
FT   CONFLICT   1038   1038       N -> I (IN REF. 1).
FT   CONFLICT   1133   1133       H -> Q (IN REF. 1).
FT   CONFLICT   1310   1310       V -> C (IN REF. 1).
FT   CONFLICT   1333   1333       G -> V (IN REF. 1).
FT   CONFLICT   1384   1384       P -> R (IN REF. 1).
FT   CONFLICT   1582   1582       G -> E (IN REF. 1).
FT   CONFLICT   1677   1682       KRRADG -> E (IN REF. 1).
FT   CONFLICT   1700   1700       C -> S (IN REF. 1).
FT   CONFLICT   1755   1755       F -> K (IN REF. 1).
FT   CONFLICT   1787   1787       T -> S (IN REF. 1).
FT   CONFLICT   1801   1822       GAIAGRVDLYEPDAFAKRPTIG -> APSPGGLICMSRCIC
FT                                ETPDNR (IN REF. 1).
FT   CONFLICT   1915   1916       LG -> PK (IN REF. 1).
FT   CONFLICT   2075   2075       R -> C (IN REF. 1).
FT   CONFLICT   2079   2079       A -> V (IN REF. 1).
FT   CONFLICT   2141   2145       ARLTP -> LRDSLN (IN REF. 1).
FT   CONFLICT   2445   2445       E -> Q (IN REF. 1).
FT   CONFLICT   2485   2489       ATDLF -> RQICS (IN REF. 1).
FT   CONFLICT   2546   2566       TVHQLFEETVQRHKDRPAVTY -> DGCISYSKRLSSATKT
FT                                ARLSHT (IN REF. 1).
FT   CONFLICT   2608   2615       MSAAVLGV -> KCPPRCSAS (IN REF. 1).
FT   CONFLICT   2644   2645       KL -> NV (IN REF. 1).
FT   CONFLICT   2713   2713       D -> H (IN REF. 1).
FT   CONFLICT   2723   2723       D -> H (IN REF. 1).
FT   CONFLICT   2876   2881       GELCVA -> RALRG (IN REF. 1).
FT   CONFLICT   2899   2900       RF -> L (IN REF. 1).
FT   CONFLICT   2958   2960       EDR -> QDA (IN REF. 1).
FT   CONFLICT   2964   2964       R -> A (IN REF. 1).
SQ   SEQUENCE   3587 AA;  401248 MW;  6B0B05A9FF32054D CRC64;

  Query Match           49.4%;  Score 41;  DB 1;  Length 3587;
  Best Local Similarity  43.8%;  Pred. No. 62;
  Matches    7;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              :| |: : | :| |||
Db       2687 AVTAENLAYMIYTSGT 2702


RESULT 3
SRF1_BACSU
ID   SRF1_BACSU     STANDARD;      PRT;  3588 AA.
AC   P27206;
DT   01-AUG-1992 (Rel. 23, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Surfactin synthetase subunit 1.
GN   SRFAA OR SRFA1 OR SRFA OR BSU03480.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=93181186; PubMed=8441623;
RA   Fuma S., Fujishima Y., Corbell N., D'Souza C., Nakano M.M.,
RA   Zuber P., Yamane K.;
RT   "Nucleotide sequence of 5' portion of srfA that contains the region
RT   required for competence establishment in Bacillus subtilis.";
RL   Nucleic Acids Res. 21:93-97(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / JH642;
RX   MEDLINE=93360813; PubMed=8355609;
RA   Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,
RA   Venema G., van Sinderen D.;
RT   "Sequence and analysis of the genetic locus responsible for surfactin
RT   synthesis in Bacillus subtilis.";
RL   Mol. Microbiol. 8:821-831(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=97124189; PubMed=8969502;
RA   Yamane K., Kumano M., Kurita K.;
RT   "The 25 degrees-36 degrees region of the Bacillus subtilis
RT   chromosome: determination of the sequence of a 146 kb segment and
RT   identification of 113 genes.";
RL   Microbiology 142:3047-3056(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [5]
RP   SEQUENCE OF 1-460 FROM N.A.
RX   MEDLINE=91154134; PubMed=1847909;
```

```
RA   Nakano M.M., Magnuson R., Myers A.M., Curry J., Grossman A.D.,
RA   Zuber P.;
RT   "srfA is an operon required for surfactin production, competence
RT   development, and efficient sporulation in Bacillus subtilis.";
RL   J. Bacteriol. 173:1770-1778(1991).
RN   [6]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=95219080; PubMed=7704255;
RA   Fujishima Y., Yamane K.;
RT   "A 10 kb nucleotide sequence at the 5' flanking region (32 degrees)
RT   of srfAA of the Bacillus subtilis chromosome.";
RL   Microbiology 141:277-279(1995).
RN   [7]
RP   SEQUENCE OF 1-38 FROM N.A.
RX   MEDLINE=91358326; PubMed=1715856;
RA   Nakano M.M., Xia L., Zuber P.;
RT   "Transcription initiation region of the srfA operon, which is
RT   controlled by the comP-comA signal transduction system in Bacillus
RT   subtilis.";
RL   J. Bacteriol. 173:5487-5493(1991).
CC   -!- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
CC       ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
CC       ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
CC   -!- COFACTOR: Contains 3 covalently bound phosphopantetheines.
CC   -!- PATHWAY: Cyclic peptide antibiotic surfactin biosynthesis.
CC   -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC       family.
CC   -!- SIMILARITY: Contains 3 acyl carrier domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13262; BAA02522.1; -.
DR   EMBL; X70356; CAA49816.1; -.
DR   EMBL; D50453; BAA08982.1; -.
DR   EMBL; Z99105; CAB12142.1; -.
DR   EMBL; M59939; AAA22815.1; -.
DR   EMBL; D30762; BAA21034.1; -.
DR   EMBL; M64702; AAA22816.1; -.
DR   PIR; I40485; I40485.
DR   HSSP; P14687; 1AMU.
DR   SubtiList; BG10168; srfAA.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_S.
DR   Pfam; PF00501; AMP-binding; 3.
DR   Pfam; PF00668; Condensation; 4.
DR   Pfam; PF00550; pp-binding; 3.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00455; AMP_BINDING; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Ligase; Antibiotic biosynthesis; Phosphopantetheine; Sporulation;
KW   Multifunctional enzyme; Repeat; Complete proteome.
FT   REPEAT        ?   1047       DOMAIN 1 (GLU-ACTIVATING).
FT   REPEAT        ?   2084       DOMAIN 2 (LEU-ACTIVATING).
FT   REPEAT        ?   3115       DOMAIN 3 (D-LEU-ACTIVATING).
FT   DOMAIN      976   1043       ACYL CARRIER (ACP) 1.
FT   DOMAIN     2013   2080       ACYL CARRIER (ACP) 2.
FT   DOMAIN     3044   3110       ACYL CARRIER (ACP) 3.
FT   BINDING    1006   1006       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2043   2043       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    3074   3074       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    146    146       M -> I (IN REF. 1 AND 5).
FT   CONFLICT    151    151       V -> L (IN REF. 1 AND 5).
FT   CONFLICT    165    165       G -> A (IN REF. 1).
FT   CONFLICT    281    281       Q -> T (IN REF. 1 AND 5).
FT   CONFLICT    460    460       D -> T (IN REF. 5).
FT   CONFLICT    540    540       A -> P (IN REF. 1).
FT   CONFLICT    562    562       Y -> I (IN REF. 1).
FT   CONFLICT    639    640       PT -> GS (IN REF. 1).
FT   CONFLICT    644    644       P -> R (IN REF. 1).
FT   CONFLICT    647    649       FLP -> LLA (IN REF. 1).
FT   CONFLICT   1026   1026       Q -> L (IN REF. 1).
FT   CONFLICT   1065   1114       HWHSSEHISSASSRMRESAIHASSSNSEGFRYSKAGARISG
FT                                INPTPRVIE -> SLAQQRTYIVSQFEDAGVGYNMPAAAIL
FT                                EGPLDIQKLERAFQGLIRRHESLR (IN REF. 1).
FT   CONFLICT   1130   1131       VC -> DS (IN REF. 1).
FT   CONFLICT   1162   1162       A -> R (IN REF. 1).
FT   CONFLICT   1456   1456       L -> V (IN REF. 1).
FT   CONFLICT   1848   1848       T -> S (IN REF. 1).
FT   CONFLICT   1892   1892       T -> D (IN REF. 1).
FT   CONFLICT   1909   1909       A -> R (IN REF. 1).
FT   CONFLICT   1971   1978       PAVFIQMD -> LRCLSKWT (IN REF. 1).
FT   CONFLICT   2050   2050       P -> L (IN REF. 1).
FT   CONFLICT   2216   2216       Q -> H (IN REF. 1).
FT   CONFLICT   2263   2264       HR -> QQ (IN REF. 1).
FT   CONFLICT   2289   2289       L -> V (IN REF. 1).
FT   CONFLICT   2347   2347       C -> S (IN REF. 1).
FT   CONFLICT   2426   2426       S -> R (IN REF. 1).
FT   CONFLICT   2609   2610       AV -> RC (IN REF. 1).
FT   CONFLICT   2684   2686       MISSING (IN REF. 1).
FT   CONFLICT   2756   2756       S -> T (IN REF. 1).
FT   CONFLICT   2897   2898       TA -> SP (IN REF. 1).
FT   CONFLICT   3026   3026       N -> P (IN REF. 1).
FT   CONFLICT   3097   3097       N -> F (IN REF. 1).
FT   CONFLICT   3272   3272       S -> A (IN REF. 1).
FT   CONFLICT   3317   3317       S -> R (IN REF. 1).
FT   CONFLICT   3452   3452       S -> Y (IN REF. 1).
FT   CONFLICT   3484   3496       HQMSHPPFFTTSH -> DEMSDAGLFTRSE (IN REF.
FT                                1).
FT   CONFLICT   3500   3501       PH -> GQ (IN REF. 1).
FT   CONFLICT   3508   3508       E -> R (IN REF. 1).
SQ   SEQUENCE   3588 AA;  402072 MW;  BC02FB157D7F1FDB CRC64;

  Query Match             49.4%;  Score 41;  DB 1;  Length 3588;
  Best Local Similarity   43.8%;  Pred. No. 62;
  Matches    7;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              :| |: : | :| |||
Db       2688 AVTAENLAYMIYTSGT 2703


RESULT 4
ENGC_BACHD
ID   ENGC_BACHD     STANDARD;      PRT;   294 AA.
AC   Q9K9Z1;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Probable GTPase engC (EC 3.6.1.-).
GN   ENGC OR BH2503.
OS   Bacillus halodurans.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=86665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RX   MEDLINE=20512582; PubMed=11058132;
RA   Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA   Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA   Horikoshi K.;
RT   "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT   halodurans and genomic sequence comparison with Bacillus subtilis.";
RL   Nucleic Acids Res. 28:4317-4331(2000).
CC   -!- FUNCTION: Unusual circulary permuted GTPase that catalyzes rapid
CC       hydrolysis of GTP with a slow catalytic turnover (By similarity).
CC   -!- SUBUNIT: Monomer (Probable).
```

```
CC   -!- SIMILARITY: Contains 1 engC GTPase domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP001515; BAB06222.1; -.
DR   PIR; G83962; G83962.
DR   HAMAP; MF_01820; -; 1.
DR   InterPro; IPR004881; DUF258.
DR   Pfam; PF03193; DUF258; 1.
DR   TIGRFAMs; TIGR00157; TIGR00157; 1.
DR   PROSITE; PS50936; ENGC_GTPASE; 1.
KW   Hydrolase; GTP-binding; Complete proteome.
FT   DOMAIN       72    221       ENGC GTPASE.
FT   NP_BIND     112    115       GTP (PROBABLE).
FT   NP_BIND     166    173       GTP (PROBABLE).
FT   NP_BIND     216    220       GTP (PROBABLE).
FT   SITE        247    260       KNUCKLE-LIKE CYSTEINE CLUSTER.
SQ   SEQUENCE   294 AA;  33449 MW;  7DC80C97AFA93A9B CRC64;

  Query Match             48.2%;  Score 40;  DB 1;  Length 294;
  Best Local Similarity   46.2%;  Pred. No. 8.2;
  Matches    6;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          5 KQMTYKVYMSGTV 17
              :|: |||||:: |:
Db        133 EQLGYKVYLTSTI 145


RESULT 5
SM2A_SCHGR
ID   SM2A_SCHGR     STANDARD;      PRT;   697 AA.
AC   Q9XZC8;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Semaphorin 2A precursor (Sema 2A) (Sema II).
GN   SEMA-2A.
OS   Schistocerca gregaria (Desert locust).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Orthopteroidea; Orthoptera; Caelifera; Acridomorpha;
OC   Acridoidea; Acrididae; Cyrtacanthacridinae; Schistocerca.
OX   NCBI_TaxID=7010;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=99203450; PubMed=10101134;
RA   Isbister C.M., Tsai A., Wong S.T., Kolodkin A.L., O'Connor T.P.;
RT   "Discrete roles for secreted and transmembrane semaphorins in neuronal
RT   growth cone guidance in vivo.";
RL   Development 126:2007-2019(1999).
CC   -!- FUNCTION: ACTS AS CHEMOREPULSIVE GUIDANCE MOLECULE CRITICAL FOR
CC       AXON FASCICULATION AND FOR DETERMINING BOTH THE INITIAL DIRECTION
CC       AND SUBSEQUENT PATHFINDING EVENTS OF THE TI AXON PROJECTION.
CC   -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN A GRADIENT IN THE DEVELOPING LIMB
CC       BUD EPITHELIUM DURING TI PIONEER AXON OUTGROWTH.
CC   -!- SIMILARITY: Belongs to the semaphorin family.
CC   -!- SIMILARITY: Contains 1 Sema domain.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF134904; AAD30114.1; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR001627; Sema.
DR   Pfam; PF01403; Sema; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00630; Sema; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Signal; Immunoglobulin domain; Neurogenesis; Developmental protein;
KW   Glycoprotein.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    697       SEMAPHORIN 2A.
FT   DOMAIN      231    515       SEMA.
FT   DOMAIN      526    634       IG-LIKE C2-TYPE.
FT   DISULFID    618    654       BY SIMILARITY.
FT   CARBOHYD     63     63       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     66     66       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    132    132       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    198    198       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    283    283       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    369    369       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    534    534       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    679    679       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   697 AA;  78673 MW;  7FE55AB4A965E1E4 CRC64;

  Query Match             48.2%;  Score 40;  DB 1;  Length 697;
  Best Local Similarity   43.8%;  Pred. No. 19;
  Matches    7;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              ::: | | | || :||
Db        419 NLVGKNMEYIVYYAGT 434


RESULT 6
OL7A_MOUSE
ID   OL7A_MOUSE     STANDARD;      PRT;   161 AA.
AC   P34985;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Olfactory receptor 7A (K18) (Fragment).
GN   OLFR7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Liver;
RX   MEDLINE=93258822; PubMed=7683976;
RA   Ressler K.J., Sullivan S.L., Buck L.B.;
RT   "A zonal organization of odorant receptor gene expression in the
RT   olfactory epithelium.";
RL   Cell 73:597-609(1993).
CC   -!- FUNCTION: Putative odorant receptor.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Olfactory epithelium.
CC   -!- MISCELLANEOUS: OLFACTORY RECEPTORS 7A-7I ARE PRODUCED BY EIGHT
CC       DIFFERENT GENES WITHIN THE OLFR7 COMPLEX.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L14568; AAA39852.1; -.
```

```
DR   PIR; A40745; A40745.
DR   MGD; MGI:104712; Olfr7.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; PARTIAL.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Multigene family;
KW   Olfaction.
FT   NON_TER       1      1
FT   DOMAIN       <1     18       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     19     37       4 (POTENTIAL).
FT   DOMAIN       38     75       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     76     98       5 (POTENTIAL).
FT   DOMAIN       99    115       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    116    139       6 (POTENTIAL).
FT   DOMAIN      140    151       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    152   >161       7 (POTENTIAL).
FT   NON_TER     161    161
SQ   SEQUENCE   161 AA;  17562 MW;  7A5140BB1EFB7FB7 CRC64;

  Query Match             47.0%;  Score 39;  DB 1;  Length 161;
  Best Local Similarity   66.7%;  Pred. No. 7;
  Matches    8;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 IAKQMTYKVYMS 14
              | | :|||| ||
Db          5 ICKPLTYKVIMS 16


RESULT 7
CBP4_YEAST
ID   CBP4_YEAST     STANDARD;      PRT;   170 AA.
AC   P37267;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   CBP4 protein, mitochondrial precursor.
GN   CBP4 OR YGR174C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94342301; PubMed=8063753;
RA   Crivellone M.D.;
RT   "Characterization of CBP4, a new gene essential for the expression of
RT   ubiquinol-cytochrome c reductase in Saccharomyces cerevisiae.";
RL   J. Biol. Chem. 269:21284-21292(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Hebling U., Hofmann B., Delius H.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Essential for the assembly and/or stability of
CC       ubiquinol-cytochrome c reductase. It has a direct effect on the
CC       correct occurrence of the Rieske protein, core 4, core 5 and
CC       apocytochrome b; it may either be involved in post-translational
CC       modification of the subunits or in the assembly of the enzyme.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL. ASSOCIATED WITH THE INNER
CC       MEMBRANE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U10700; AAA61566.1; -.
DR   EMBL; Z72959; CAA97200.1; -.
DR   PIR; S64488; S64488.
DR   GermOnline; 141486; -.
DR   SGD; S0003406; CBP4.
KW   Mitochondrion; Transit peptide.
FT   TRANSIT       1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?    170       CBP4 PROTEIN.
FT   CONFLICT     65     65       S -> F (IN REF. 1).
SQ   SEQUENCE   170 AA;  20219 MW;  D88F92EADF0B366E CRC64;

  Query Match             47.0%;  Score 39;  DB 1;  Length 170;
  Best Local Similarity   61.5%;  Pred. No. 7.3;
  Matches    8;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYM 13
              :|||||  || |:
Db         10 AVIAKQRQYKHYL 22


RESULT 8
PLSX_AQUAE
ID   PLSX_AQUAE     STANDARD;      PRT;   337 AA.
AC   O67186;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Fatty acid/phospholipid synthesis protein plsX.
GN   PLSX OR AQ_1101.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- FUNCTION: Not known, probably involved in fatty acid or
CC       phospholipid synthesis (By similarity).
CC   -!- SIMILARITY: Belongs to the plsX family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000723; AAC07145.1; -.
DR   PIR; G70394; G70394.
DR   HAMAP; MF_00019; -; 1.
DR   InterPro; IPR003664; FA_synthesis.
DR   Pfam; PF02504; FA_synthesis; 1.
DR   ProDom; PD006974; FA_synthesis; 1.
DR   TIGRFAMs; TIGR00182; plsX; 1.
KW   Fatty acid biosynthesis; Phospholipid biosynthesis; Complete proteome.
SQ   SEQUENCE   337 AA;  36266 MW;  C6E51574FA15D508 CRC64;

  Query Match             47.0%;  Score 39;  DB 1;  Length 337;
  Best Local Similarity   50.0%;  Pred. No. 14;
  Matches    6;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 AKQMTYKVYMSG 15
              ||:: ||:|: |
Db         26 AKELGYKIYLVG 37


RESULT 9
SERC_NEIMA
ID   SERC_NEIMA     STANDARD;      PRT;   368 AA.
AC   O34370; O33382; O33383; O33384; O33386;
```

```
DT   15-DEC-1998 (Rel. 37, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Phosphoserine aminotransferase (EC 2.6.1.52) (PSAT).
GN   SERC OR NMA1894.
OS   Neisseria meningitidis (serogroup A).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC   Neisseriaceae; Neisseria.
OX   NCBI_TaxID=65699;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Various strains;
RX   MEDLINE=98010345; PubMed=9350862;
RA   Morelli G., Malorny B., Mueller K., Seiler A., Wang J.-F.,
RA   del Valle J., Achtman M.;
RT   "Clonal descent and microevolution of Neisseria meningitidis during
RT   30 years of epidemic spread.";
RL   Mol. Microbiol. 25:1047-1064(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX   MEDLINE=20222556; PubMed=10761919;
RA   Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA   Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA   Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA   Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA   Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA   Whitehead S., Spratt B.G., Barrell B.G.;
RT   "Complete DNA sequence of a serogroup A strain of Neisseria
RT   meningitidis Z2491.";
RL   Nature 404:502-506(2000).
CC   -!- CATALYTIC ACTIVITY: O-phospho-L-serine + 2-oxoglutarate = 3-
CC       phosphonooxypyruvate + L-glutamate.
CC   -!- COFACTOR: Pyridoxal phosphate.
CC   -!- PATHWAY: Required both in major phosphorylated pathway of serine
CC       biosynthesis and in the biosynthesis of pyridoxine.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to class-V of pyridoxal-phosphate-dependent
CC       aminotransferases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF004820; AAC32675.1; -.
DR   EMBL; AF004821; AAC32679.1; -.
DR   EMBL; AF004822; AAC32683.1; -.
DR   EMBL; AF004823; AAC32687.1; -.
DR   EMBL; AF004824; AAC32691.1; -.
DR   EMBL; AF004825; AAC32695.1; -.
DR   EMBL; AF004826; AAC32699.1; -.
DR   EMBL; AL162757; CAB85115.1; -.
DR   PIR; F81816; F81816.
DR   HSSP; P23721; 1BJN.
DR   HAMAP; MF_00160; -; 1.
DR   InterPro; IPR000192; Aminotrans_V.
DR   InterPro; IPR003248; Pser_amintransf.
DR   Pfam; PF00266; aminotran_5; 1.
DR   ProDom; PD001544; Pser_amintransf; 1.
DR   TIGRFAMs; TIGR01364; serC_1; 1.
DR   PROSITE; PS00595; AA_TRANSFER_CLASS_5; 1.
KW   Serine biosynthesis; Pyridoxine biosynthesis; Transferase;
KW   Aminotransferase; Pyridoxal phosphate; Complete proteome.
FT   BINDING     203    203       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   VARIANT     168    168       R -> C (IN STRAINS B293, Z3910 AND
FT                                Z3918).
FT   VARIANT     192    192       A -> S (IN STRAIN Z3524).
FT   VARIANT     237    237       I -> L (IN STRAINS B293, Z3524, Z3910,
FT                                Z3915 AND Z3918).
FT   VARIANT     240    240       D -> E (IN STRAINS Z3915 AND Z3524).
FT   VARIANT     289    289       G -> D (IN STRAINS B293, Z3910 AND
FT                                Z3918).
FT   VARIANT     336    336       T -> S (IN STRAIN Z4296).
SQ   SEQUENCE   368 AA;  41388 MW;  3D3E305853698537 CRC64;

  Query Match             47.0%;  Score 39;  DB 1;  Length 368;
  Best Local Similarity   70.0%;  Pred. No. 16;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          8 TYKVYMSGTV 17
              || :|||| |
Db        248 TYAIYMSGLV 257


RESULT 10
SERC_NEIMB
ID   SERC_NEIMB     STANDARD;      PRT;   368 AA.
AC   P57007;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Phosphoserine aminotransferase (EC 2.6.1.52) (PSAT).
GN   SERC OR NMB1640.
OS   Neisseria meningitidis (serogroup B).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC   Neisseriaceae; Neisseria.
OX   NCBI_TaxID=491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / Serogroup B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
CC   -!- CATALYTIC ACTIVITY: O-phospho-L-serine + 2-oxoglutarate = 3-
CC       phosphonooxypyruvate + L-glutamate.
CC   -!- COFACTOR: Pyridoxal phosphate.
CC   -!- PATHWAY: Required both in major phosphorylated pathway of serine
CC       biosynthesis and in the biosynthesis of pyridoxine.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to class-V of pyridoxal-phosphate-dependent
CC       aminotransferases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE002514; AAF41989.1; -.
DR   PIR; H81059; H81059.
DR   HSSP; P23721; 1BJN.
DR   TIGR; NMB1640; -.
DR   HAMAP; MF_00160; -; 1.
DR   InterPro; IPR000192; Aminotrans_V.
DR   InterPro; IPR003248; Pser_amintransf.
DR   Pfam; PF00266; aminotran_5; 1.
DR   ProDom; PD001544; Pser_amintransf; 1.
DR   TIGRFAMs; TIGR01364; serC_1; 1.
DR   PROSITE; PS00595; AA_TRANSFER_CLASS_5; 1.
KW   Serine biosynthesis; Pyridoxine biosynthesis; Transferase;
KW   Aminotransferase; Pyridoxal phosphate; Complete proteome.
```

```
FT   BINDING     203    203       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   368 AA;  41393 MW;  97DFCE52BBE5E021 CRC64;

  Query Match             47.0%;  Score 39;  DB 1;  Length 368;
  Best Local Similarity   70.0%;  Pred. No. 16;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          8 TYKVYMSGTV 17
              || :|||| |
Db        248 TYAIYMSGLV 257


RESULT 11
SP98_YEAST
ID   SP98_YEAST     STANDARD;      PRT;   846 AA.
AC   P53540;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Spindle pole body component SPC98.
GN   SPC98 OR YNL126W OR N1222 OR N1879.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=96109932; PubMed=8619318;
RA   Mallet L., Bussereau F., Jacquet M.;
RT   "A 43.5 kb segment of yeast chromosome XIV, which contains MFA2,
RT   MEP2, CAP/SRV2, NAM9, FKB1/FPR1/RBP1, MOM22 and CPT1, predicts an
RT   adenosine deaminase gene and 14 new open reading frames.";
RL   Yeast 11:1195-1209(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97245296; PubMed=9090055;
RA   de Antoni A., D'Angelo M., Dal Pero F., Sartorello F., Pandolfo D.,
RA   Pallavicini A., Lanfranchi G., Valle G.;
RT   "The DNA sequence of cosmid 14-13b from chromosome XIV of
RT   Saccharomyces cerevisiae reveals an unusually high number of
RT   overlapping open reading frames.";
RL   Yeast 13:261-266(1997).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=96324398; PubMed=8670895;
RA   Geissler S., Pereira G., Spang A., Knop M., Soues S., Kilmartin J.V.,
RA   Schiebel E.;
RT   "The spindle pole body component Spc98p interacts with the
RT   gamma-tubulin-like Tub4p of Saccharomyces cerevisiae at the sites of
RT   microtubule attachment.";
RL   EMBO J. 15:3899-3911(1996).
CC   -!- FUNCTION: Involved in microtubule organization by the microtubule
CC       organizing centre, the spindle pole body (SPB). Probably part of
CC       the microtubule attachment site at the SPB.
CC   -!- SUBUNIT: Interacts with TUB4 and SPC97.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Belongs to the GCP family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z46843; CAA86899.1; -.
DR   EMBL; Z69382; CAA93378.1; -.
DR   EMBL; Z71402; CAA96007.1; -.
DR   PIR; S59262; S59262.
DR   GermOnline; 143132; -.
DR   SGD; S0005070; SPC98.
DR   GO; GO:0005822; C:inner plaque of spindle pole body; IDA.
DR   GO; GO:0005824; C:outer plaque of spindle pole body; IDA.
DR   GO; GO:0005200; F:structural constituent of cytoskeleton; IPI.
DR   GO; GO:0007020; P:microtubule nucleation; IPI.
DR   GO; GO:0000071; P:mitotic spindle assembly (sensu Saccharomyces); IMP.
DR   InterPro; IPR007259; Spc97_Spc98.
DR   Pfam; PF04130; Spc97_Spc98; 1.
KW   Microtubule; Nuclear protein.
SQ   SEQUENCE   846 AA;  98226 MW;  803048B05D5E5105 CRC64;

  Query Match             47.0%;  Score 39;  DB 1;  Length 846;
  Best Local Similarity   46.2%;  Pred. No. 35;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 IAKQMTYKVYMSG 15
              | |:: ||::| |
Db        394 IPKELAYKIFMIG 406


RESULT 12
ACVS_CEPAC
ID   ACVS_CEPAC     STANDARD;      PRT;  3712 AA.
AC   P25464;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   N-(5-amino-5-carboxypentanoyl)-L-cysteinyl-D-valine synthase
DE   (EC 6.3.2.26) (Delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine
DE   synthetase) (ACV synthetase) (ACVS).
GN   PCBAB.
OS   Cephalosporium acremonium (Acremonium chrysogenum).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreomycetidae; Hypocreales; Hypocreaceae; mitosporic Hypocreaceae;
OC   Acremonium.
OX   NCBI_TaxID=5044;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91177827; PubMed=1706706;
RA   Gutierrez S., Diez B., Montenegro E., Martin J.F.;
RT   "Characterization of the Cephalosporium acremonium pcbAB gene
RT   encoding alpha-aminoadipyl-cysteinyl-valine synthetase, a large
RT   multidomain peptide synthetase: linkage to the pcbC gene as a cluster
RT   of early cephalosporin biosynthetic genes and evidence of multiple
RT   functional domains.";
RL   J. Bacteriol. 173:2354-2365(1991).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=ATCC 11550;
RX   MEDLINE=91168300; PubMed=2076552;
RA   Hoskins J.A., O'Callaghan N., Queener S.W., Cantwell C.A., Wood J.S.,
RA   Chen V.J., Skatrud P.L.;
RT   "Gene disruption of the pcbAB gene encoding ACV synthetase in
RT   Cephalosporium acremonium.";
RL   Curr. Genet. 18:523-530(1990).
CC   -!- FUNCTION: Each of the constituent amino acids of the tripeptide
CC       acv are activated as aminoacyl-adenylates with peptide bonds
CC       formed through the participation of amino acid thiolester
CC       intermediates.
CC   -!- CATALYTIC ACTIVITY: L-2-aminohexanedioate + L-cysteine + L-valine
CC       + 3 ATP = N-[L-5-amino-5-carboxypentanoyl]-L-cysteinyl-D-valine +
CC       3 AMP + 3 diphosphate.
CC   -!- COFACTOR: Contains 3 covalently bound phosphopantetheines
CC       (Potential).
CC   -!- PATHWAY: Biosynthesis of penicillin and cephalosporin; first step.
CC   -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC       family.
CC   -!- SIMILARITY: Contains 3 acyl carrier domains.
DR   PIR; A38531; YGCEVC.
DR   HSSP; P14687; 1AMU.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_S.
```

```
DR   InterPro; IPR000379; Ser_estrs.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00501; AMP-binding; 3.
DR   Pfam; PF00668; Condensation; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00455; AMP_BINDING; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Ligase; Antibiotic biosynthesis; Multifunctional enzyme;
KW   Repeat; Phosphopantetheine.
FT   REPEAT      234   1062       DOMAIN 1 (ADIPATE-ACTIVATING).
FT   REPEAT     1335   2162       DOMAIN 2 (CYSTEINE-ACTIVATING).
FT   REPEAT     2409   3387       DOMAIN 3 (VALINE-ACTIVATING).
FT   DOMAIN      795    864       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1880   1953       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2960   3027       ACYL CARRIER (ACP) 3.
FT   BINDING     827    827       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1916   1916       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    2990   2990       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   3568   3568       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   3712 AA;  414767 MW;  4EE3C1EB5EBEF9B7 CRC64;

  Query Match             47.0%;  Score 39;  DB 1;  Length 3712;
  Best Local Similarity   53.8%;  Pred. No. 1.5e+02;
  Matches    7;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          4 AKQMTYKVYMSGT 16
              :||: |  | |||
Db        414 SKQLAYVTYTSGT 426


RESULT 13
YF62_ARCFU
ID   YF62_ARCFU     STANDARD;      PRT;   140 AA.
AC   O28710;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein AF1562.
GN   AF1562.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC   Archaeoglobaceae; Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000994; AAB89687.1; -.
DR   PIR; A69445; A69445.
DR   TIGR; AF1562; -.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     20     42       POTENTIAL.
FT   TRANSMEM     88    110       POTENTIAL.
FT   TRANSMEM    115    137       POTENTIAL.
SQ   SEQUENCE   140 AA;  15667 MW;  937DCB5585A17991 CRC64;

  Query Match             45.8%;  Score 38;  DB 1;  Length 140;
  Best Local Similarity   37.5%;  Pred. No. 9.3;
  Matches    6;  Conservative    6;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 VIAKQMTYKVYMSGTV 17
              :|   :|: :|:|||:
Db         29 IIFMAITFAIYVSGTL 44


RESULT 14
YAJG_ECOLI
ID   YAJG_ECOLI     STANDARD;      PRT;   192 AA.
AC   P36671; P77210;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical lipoprotein yajG precursor.
GN   YAJG OR B0434 OR C0546.
OS   Escherichia coli, and
OS   Escherichia coli O6.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562, 217992;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94049112; PubMed=8231804;
RA   Lindquist S., Weston-Hafer K., Schmidt H., Pul C., Korfmann G.,
RA   Erickson J., Sanders C., Martin H.H., Normark S.;
RT   "AmpG, a signal transducer in chromosomal beta-lactamase induction.";
RL   Mol. Microbiol. 9:703-715(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA   Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA   Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC       (Probable).
CC   -!- SIMILARITY: TO H.INFLUENZAE HI0162.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; S67816; AAB28883.2; -.
DR   EMBL; AE000149; AAC73537.1; ALT_INIT.
DR   EMBL; U82664; AAB40190.1; ALT_INIT.
DR   EMBL; AE016756; AAN79024.1; ALT_INIT.
DR   EcoGene; EG12182; yajG.
DR   InterPro; IPR005619; Lipoprotein_16.
DR   InterPro; IPR000437; Prok_lipoprot_S.
DR   Pfam; PF03923; Lipoprotein_16; 1.
DR   ProDom; PD036382; Lipoprotein_16; 1.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Hypothetical protein; Membrane; Lipoprotein; Signal;
KW   Complete proteome; Palmitate.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    192       HYPOTHETICAL LIPOPROTEIN YAJG.
FT   LIPID        18     18       N-palmitoyl cysteine (Potential).
FT   LIPID        18     18       S-diacylglycerol cysteine (Potential).
SQ   SEQUENCE   192 AA;  20950 MW;  9E9E6568E9253451 CRC64;

  Query Match             45.8%;  Score 38;  DB 1;  Length 192;
  Best Local Similarity   57.1%;  Pred. No. 13;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 VIAKQMTYKVYMSG 15
              |: |||| : || |
Db         83 VLEKQMTARGYMVG 96


RESULT 15
BCHL_RHOCA
ID   BCHL_RHOCA     STANDARD;      PRT;   304 AA.
AC   P26237;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Light-independent protochlorophyllide reductase iron-sulfur ATP-
DE   binding protein (EC 1.18.-.-) (LI-POR subunit L) (DPOR subunit L).
GN   BCHL.
OS   Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC   Rhodobacteraceae; Rhodobacter.
OX   NCBI_TaxID=1061;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90368552; PubMed=2203738;
RA   Yang Z., Bauer C.E.;
RT   "Rhodobacter capsulatus genes involved in early steps of the
RT   bacteriochlorophyll biosynthetic pathway.";
RL   J. Bacteriol. 172:5001-5010(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SB1003 / St Louis;
RA   Burke D.H., Alberti M., Armstrong G.A., Hearst J.E.;
RL   Submitted (NOV-1991) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   PRELIMINARY SEQUENCE FROM N.A.
RX   MEDLINE=84259352; PubMed=6744416;
RA   Youvan D.C., Bylina E.J., Alberti M., Begusch H., Hearst J.E.;
RT   "Nucleotide and deduced polypeptide sequences of the photosynthetic
RT   reaction-center, B870 antenna, and flanking polypeptides from R.
RT   capsulata.";
RL   Cell 37:949-957(1984).
RN   [4]
RP   CHARACTERIZATION.
RC   STRAIN=SB1003 / CB1029;
RX   MEDLINE=20378986; PubMed=10811655;
RA   Fujita Y., Bauer C.E.;
RT   "Reconstitution of light-independent protochlorophyllide reductase
RT   from purified bchL and bchN-bchB subunits. In vitro confirmation of
RT   nitrogenase-like features of a bacteriochlorophyll biosynthesis
RT   enzyme.";
RL   J. Biol. Chem. 275:23583-23588(2000).
RN   [5]
RP   CHARACTERIZATION.
RA   Fujita Y.;
RL   Unpublished observations (JUL-2001).
CC   -!- FUNCTION: Uses Mg-ATP and reduced ferredoxin to reduce ring D of
CC       protochlorophyllide (Pchlide) to form chlorophyllide a (Chlide).
CC       This reaction is light-independent.
CC   -!- PATHWAY: Light-independent bacteriochlorophyll biosynthesis.
CC   -!- SUBUNIT: Protochlorophyllide reductase is thought to be composed
CC       of three subunits; bchL, bchN and bchB. Homodimer of bchL subunit
CC       (By similarity).
CC   -!- SIMILARITY: Belongs to the nifH / bchL / chlL family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M34843; AAA26098.1; -.
DR   EMBL; Z11165; CAA77523.1; -.
DR   EMBL; K01183; -; NOT_ANNOTATED_CDS.
DR   PIR; B36716; B36716.
DR   PIR; H28771; H28771.
DR   HSSP; P00456; 1CP2.
DR   HAMAP; MF_00355; -; 1.
DR   InterPro; IPR000392; NitrogenaseII.
DR   InterPro; IPR005971; Protochl_reductF.
DR   Pfam; PF00142; fer4_NifH; 1.
DR   PRINTS; PR00091; NITROGNASEII.
DR   TIGRFAMs; TIGR01281; DPOR_bchL; 1.
DR   PROSITE; PS00746; NIFH_FRXC_1; 1.
DR   PROSITE; PS00692; NIFH_FRXC_2; 1.
KW   Oxidoreductase; Photosynthesis; Bacteriochlorophyll biosynthesis;
KW   ATP-binding; Iron-sulfur; 4Fe-4S.
FT   NP_BIND      43     50       ATP (POTENTIAL).
FT   METAL       131    131       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT   METAL       165    165       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ   SEQUENCE   304 AA;  33204 MW;  3A49C39BCF15AECC CRC64;

  Query Match             45.8%;  Score 38;  DB 1;  Length 304;
  Best Local Similarity   47.1%;  Pred. No. 20;
  Matches    8;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              :| || : ||| ::| |
Db        198 AVQAKSVNYKVRLAGCV 214


Search completed: August 12, 2004, 06:20:06
Job time : 3.04508 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 4.42418 Seconds
(without alignments)
1212.385 Million cell updates/sec

Title: US-09-890-463-2
Perfect score: 83
Sequence: 1 SVIAKQMTYKVYMSGTV 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	83	100.0	221	5	Q95P04	Q95p04 goniopora t
2	45	54.2	398	16	Q97LU2	Q97lu2 clostridium
3	43	51.8	389	16	Q8XIP1	Q8xip1 clostridium
4	43	51.8	424	4	Q9UF27	Q9uf27 homo sapien
5	43	51.8	4243	4	Q86WI1	Q86wi1 homo sapien
6	41	49.4	227	5	Q8MU45	Q8mu45 condylactis
7	41	49.4	227	5	Q95W11	Q95w11 condylactis
8	41	49.4	227	5	Q95W86	Q95w86 condylactis
9	41	49.4	592	17	Q96YI5	Q96yi5 sulfolobus
10	41	49.4	3583	2	Q45675	Q45675 bacillus su
11	40	48.2	218	16	Q9CI65	Q9ci65 lactococcus
12	40	48.2	225	5	Q9U6Y8	Q9u6y8 discosoma s
13	40	48.2	230	5	Q9GTJ7	Q9gtj7 discosoma s
14	40	48.2	263	16	Q97FT4	Q97ft4 clostridium
15	40	48.2	268	10	Q7XDZ1	Q7xdz1 oryza sativ
16	40	48.2	285	16	Q9KBK0	Q9kbk0 bacillus ha
17	40	48.2	294	16	Q9K9Z1	Q9k9z1 bacillus ha
18	40	48.2	557	16	Q888M2	Q888m2 pseudomonas
19	40	48.2	870	10	Q9SUC1	Q9suc1 arabidopsis
20	40	48.2	1086	16	Q8YTR6	Q8ytr6 anabaena sp
21	40	48.2	1307	10	Q9LVN1	Q9lvn1 arabidopsis
22	39	47.0	150	2	Q9R2Y3	Q9r2y3 neisseria m
23	39	47.0	150	2	O52460	O52460 neisseria m
24	39	47.0	150	2	O52461	O52461 neisseria m
25	39	47.0	150	2	O54619	O54619 neisseria m
26	39	47.0	150	2	O52462	O52462 neisseria m
27	39	47.0	150	2	O52469	O52469 neisseria m
28	39	47.0	150	2	O52463	O52463 neisseria m
29	39	47.0	150	2	O88149	O88149 neisseria m
30	39	47.0	150	2	O52464	O52464 neisseria m
31	39	47.0	150	2	Q9R3P3	Q9r3p3 neisseria m
32	39	47.0	150	2	O52465	O52465 neisseria m
33	39	47.0	150	2	O52466	O52466 neisseria m
34	39	47.0	150	2	O52467	O52467 neisseria m
35	39	47.0	150	2	O52468	O52468 neisseria m
36	39	47.0	150	2	O54532	O54532 neisseria m
37	39	47.0	227	5	Q95W85	Q95w85 radianthus
38	39	47.0	228	5	Q86LV4	Q86lv4 radianthus
39	39	47.0	277	16	O50740	O50740 borrelia bu
40	39	47.0	286	9	Q9AYV4	Q9ayv4 lactococcus
41	39	47.0	313	11	Q9EQB3	Q9eqb3 mus musculu
42	39	47.0	346	16	O84778	O84778 chlamydia t
43	39	47.0	346	16	Q9PLF0	Q9plf0 chlamydia m
44	39	47.0	368	2	Q84I19	Q84i19 neisseria g
45	39	47.0	368	2	Q84I18	Q84i18 neisseria g

ALIGNMENTS

RESULT 1
Q95P04
ID Q95P04 PRELIMINARY; PRT; 221 AA.
AC Q95P04;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GFP-like chromoprotein.
OS Goniopora tenuidens.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Fungiina; Poritidae; Goniopora.
OX NCBI_TaxID=75301;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21538626; PubMed=11682051;
RA Gurskaya N.G., Fradkov A.F., Terskikh A., Matz M.V., Labas Y.A.,
RA Martynov V.I., Yanushevich Y.G., Lukyanov K.A., Lukyanov S.A.;
RT "GFP-like chromoproteins as a source of far-red fluorescent
RT proteins(1).";
RL FEBS Lett. 507:16-20(2001).
DR EMBL; AF383156; AAL27542.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 221 AA; 24918 MW; 93F9F4B5C2003CB4 CRC64;

Query Match 100.0%; Score 83; DB 5; Length 221;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
|||||||||||||||||
Db 2 SVIAKQMTYKVYMSGTV 18

RESULT 2
Q97LU2

```
ID   Q97LU2      PRELIMINARY;      PRT;   398 AA.
AC   Q97LU2;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Protein of short-chain alcohol dehydrogenase family.
GN   CAC0462.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
DR   EMBL; AE007561; AAK78442.1; -.
DR   PIR; G96956; G96956.
KW   Complete proteome.
SQ   SEQUENCE   398 AA;  45650 MW;  59324A21CA466DFC CRC64;

  Query Match             54.2%;  Score 45;  DB 16;  Length 398;
  Best Local Similarity   47.1%;  Pred. No. 14;
  Matches    8;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              | |     |||:|  ||:
Db        224 SYIGSPRTYKIYREGTI 240


RESULT 3
Q8XIP1
ID   Q8XIP1      PRELIMINARY;      PRT;   389 AA.
AC   Q8XIP1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein CPE2074.
GN   CPE2074.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / Type A;
RX   MEDLINE=21664373; PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR   EMBL; AP003192; BAB81780.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   389 AA;  43138 MW;  36E1230CC803E7C5 CRC64;

  Query Match             51.8%;  Score 43;  DB 16;  Length 389;
  Best Local Similarity   41.2%;  Pred. No. 32;
  Matches    7;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              | |  ::|| :|  ||:
Db        223 SYIGPEVTYPIYREGTI 239


RESULT 4
Q9UF27
ID   Q9UF27      PRELIMINARY;      PRT;   424 AA.
AC   Q9UF27;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   DKFZP586C1021.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RA   Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL133640; CAB63761.1; -.
DR   PIR; T43498; T43498.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   424 AA;  46402 MW;  35523FD7C62313A2 CRC64;

  Query Match             51.8%;  Score 43;  DB 4;  Length 424;
  Best Local Similarity   43.8%;  Pred. No. 34;
  Matches    7;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              |::|   :|:|| :||
Db          1 SIVALNKSYEVYFTGT 16


RESULT 5
Q86WI1
ID   Q86WI1      PRELIMINARY;      PRT;  4243 AA.
AC   Q86WI1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Fibrocystin L.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22508206; PubMed=12620974;
RA   Hogan M.C., Griffin M.D., Rossetti S., Torres V.E., Ward C.J.,
RA   Harris P.C.;
RT   "PKHDL1, a homolog of the autosomal recessive polycystic kidney
RT   disease gene, encodes a receptor with inducible T lymphocyte
RT   expression.";
RL   Hum. Mol. Genet. 12:685-698(2003).
DR   EMBL; AY219181; AAO60072.1; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR002909; IPT_TIG.
DR   InterPro; IPR006626; PbH1.
DR   Pfam; PF01833; TIG; 14.
DR   SMART; SM00429; IPT; 14.
DR   SMART; SM00710; PbH1; 10.
SQ   SEQUENCE   4243 AA;  465745 MW;  36FE9DE63F4931E7 CRC64;

  Query Match             51.8%;  Score 43;  DB 4;  Length 4243;
  Best Local Similarity   43.8%;  Pred. No. 3.6e+02;
  Matches    7;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              |::|   :|:|| :||
Db       3820 SIVALNKSYEVYFTGT 3835


RESULT 6
Q8MU45
ID   Q8MU45      PRELIMINARY;      PRT;   227 AA.
```

```
AC   Q8MU45;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Green fluorescent protein-like protein.
OS   Condylactis gigantea (Giant anemone).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC   Nynantheae; Actiniidae; Condylactis.
OX   NCBI_TaxID=47073;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Matz M.V., Lukyanov S.A.;
RT   "Diversity and evolution of GFP-like fluorescent proteins.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY037777; AAK71343.1; -.
DR   GO; GO:0006091; P:energy pathways; IEA.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   227 AA;  25384 MW;  D3C6B02F490F3D21 CRC64;

  Query Match             49.4%;  Score 41;  DB 5;  Length 227;
  Best Local Similarity   43.8%;  Pred. No. 41;
  Matches    7;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 VIAKQMTYKVYMSGTV 17
              :: : |  |:|| |||
Db          4 LLKESMRIKIYMEGTV 19


RESULT 7
Q95W11
ID   Q95W11      PRELIMINARY;      PRT;   227 AA.
AC   Q95W11;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   GFP-like chromoprotein.
OS   Condylactis passiflora.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC   Nynantheae; Actiniidae; Condylactis.
OX   NCBI_TaxID=175772;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21538626; PubMed=11682051;
RA   Gurskaya N.G., Fradkov A.F., Terskikh A., Matz M.V., Labas Y.A.,
RA   Martynov V.I., Yanushevich Y.G., Lukyanov K.A., Lukyanov S.A.;
RT   "GFP-like chromoproteins as a source of far-red fluorescent
RT   proteins(1).";
RL   FEBS Lett. 507:16-20(2001).
DR   EMBL; AF383155; AAL27541.1; -.
DR   GO; GO:0006091; P:energy pathways; IEA.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   227 AA;  25446 MW;  E51CC017108593E3 CRC64;

  Query Match             49.4%;  Score 41;  DB 5;  Length 227;
  Best Local Similarity   43.8%;  Pred. No. 41;
  Matches    7;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 VIAKQMTYKVYMSGTV 17
              :: : |  |:|| |||
Db          4 LLKESMRIKIYMEGTV 19


RESULT 8
Q95W86
ID   Q95W86      PRELIMINARY;      PRT;   227 AA.
AC   Q95W86;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   GFP-like chromoprotein.
OS   Condylactis gigantea (Giant anemone).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC   Nynantheae; Actiniidae; Condylactis.
OX   NCBI_TaxID=47073;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21538626; PubMed=11682051;
RA   Gurskaya N.G., Fradkov A.F., Terskikh A., Matz M.V., Labas Y.A.,
RA   Martynov V.I., Yanushevich Y.G., Lukyanov K.A., Lukyanov S.A.;
RT   "GFP-like chromoproteins as a source of far-red fluorescent
RT   proteins(1).";
RL   FEBS Lett. 507:16-20(2001).
DR   EMBL; AF363775; AAL27537.1; -.
DR   GO; GO:0006091; P:energy pathways; IEA.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   227 AA;  25416 MW;  BCFA5A4CBCE1B3F7 CRC64;

  Query Match             49.4%;  Score 41;  DB 5;  Length 227;
  Best Local Similarity   43.8%;  Pred. No. 41;
  Matches    7;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 VIAKQMTYKVYMSGTV 17
              :: : |  |:|| |||
Db          4 LLKESMRIKIYMEGTV 19


RESULT 9
Q96YI5
ID   Q96YI5      PRELIMINARY;      PRT;   592 AA.
AC   Q96YI5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative glucosamine--fructose-6-phosphate aminotransferase.
GN   ST2186.
OS   Sulfolobus tokodaii.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=111955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JCM 10545 / 7;
RX   MEDLINE=21456156; PubMed=11572479;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA   Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA   Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA   Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA   Oshima T., Kikuchi H.;
RT   "Complete genome sequence of an aerobic thermoacidophilic
RT   Crenarchaeon, Sulfolobus tokodaii strain7.";
RL   DNA Res. 8:123-140(2001).
DR   EMBL; AP000989; BAB67292.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0004360; F:glutamine-fructose-6-phosphate transaminase. . .; IEA.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR   InterPro; IPR000583; GATase_2.
DR   InterPro; IPR005855; GlmS.
DR   InterPro; IPR001347; SIS.
DR   Pfam; PF00310; GATase_2; 1.
DR   Pfam; PF01380; SIS; 2.
DR   TIGRFAMs; TIGR01135; glmS; 1.
DR   PROSITE; PS00443; GATASE_TYPE_II; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   592 AA;  65796 MW;  3CED613D9A0EB7ED CRC64;
```

```
  Query Match             49.4%;  Score 41;  DB 17;  Length 592;
  Best Local Similarity   53.8%;  Pred. No. 1.1e+02;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYM 13
              | ||::  ||:||
Db        379 SAIARESDYKIYM 391


RESULT 10
Q45675
ID   Q45675      PRELIMINARY;      PRT;  3583 AA.
AC   Q45675;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Surfactin synthetase.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 trpC2;
RX   MEDLINE=95219089; PubMed=7704264;
RA   Fabret C., Quentin Y., Guiseppi A., Busuttil J., Haiech J.,
RA   Denizot F.;
RT   "Analysis of errors in finished DNA sequences: the surfactin operon of
RT   Bacillus subtilis as an example.";
RL   Microbiology 141:345-350(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 trpC2;
RX   MEDLINE=93181186; PubMed=8441623;
RA   Fuma S., Fujishima Y., Corbell N., D'Souza C., Nakano M.M., Zuber P.,
RA   Yamane K.;
RT   "Nucleotide sequence of 5' portion of srfA that contains the region
RT   required for competence establishment in Bacillus subtilis.";
RL   Nucleic Acids Res. 21:93-97(1993).
DR   EMBL; X72672; CAA51223.1; -.
DR   HSSP; P14687; 1AMU.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR006162; Ppantne_S.
DR   InterPro; IPR006163; Pp_bind.
DR   Pfam; PF00501; AMP-binding; 3.
DR   Pfam; PF00668; Condensation; 4.
DR   Pfam; PF00550; pp-binding; 3.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00455; AMP_BINDING; 3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW   Phosphopantetheine.
FT   CONFLICT    113    113       R -> A (IN REF. 2).
FT   CONFLICT    940    940       L -> R (IN REF. 2).
FT   CONFLICT   1310   1310       V -> C (IN REF. 2).
FT   CONFLICT   1782   1782       T -> S (IN REF. 2).
FT   CONFLICT   1817   1817       G -> R (IN REF. 2).
FT   CONFLICT   2070   2070       R -> C (IN REF. 2).
FT   CONFLICT   2135   2135       R -> A (IN REF. 2).
FT   CONFLICT   2390   2390       L -> A (IN REF. 2).
FT   CONFLICT   2481   2481       A -> P (IN REF. 2).
FT   CONFLICT   2486   2486       A -> L (IN REF. 2).
FT   CONFLICT   2542   2542       T -> E (IN REF. 2).
FT   CONFLICT   2544   2544       H -> A (IN REF. 2).
FT   CONFLICT   2563   2563       N -> D (IN REF. 2).
FT   CONFLICT   2603   2604       EM -> GK (IN REF. 2).
FT   CONFLICT   2641   2641       L -> P (IN REF. 2).
FT   CONFLICT   2895   2895       R -> P (IN REF. 2).
SQ   SEQUENCE   3583 AA;  400937 MW;  A257AC7643C4C64C CRC64;

  Query Match             49.4%;  Score 41;  DB 2;  Length 3583;
  Best Local Similarity   43.8%;  Pred. No. 7e+02;
  Matches    7;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              :| |: : | :| |||
Db       2683 AVTAENLAYMIYTSGT 2698


RESULT 11
Q9CI65
ID   Q9CI65      PRELIMINARY;      PRT;  218 AA.
AC   Q9CI65;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein yfaA.
GN   YFAA OR LL0501.
OS   Lactococcus lactis (subsp. lactis) (Streptococcus lactis), and
OS   Lactococcus lactis.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX   NCBI_TaxID=1360, 1358;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=L.lactis (subsp. lactis); STRAIN=IL1403;
RX   MEDLINE=21235186; PubMed=11337471;
RA   Bolotin A., Wincker P., Mauger S., Jaillon O., Malarme K.,
RA   Weissenbach J., Ehrlich S.D., Sorokin A.;
RT   "The complete genome sequence of the lactic acid bacterium Lactococcus
RT   lactis ssp. lactis IL1403.";
RL   Genome Res. 11:731-753(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=L.lactis; STRAIN=IL1403;
RC   TRANSPOSON=transposon-like element TnX;
RX   MEDLINE=20000172; PubMed=10532372;
RA   Bolotin A., Mauger S., Malarme K., Ehrlich S.D., Sorokin A.;
RT   "Low-redundancy sequencing of the entire Lactococcus lactis IL1403
RT   genome.";
RL   Antonie Van Leeuwenhoek 76:27-76(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=L.lactis; STRAIN=IL1403;
RC   TRANSPOSON=transposon-like element TnX;
RA   Calero S., Ehrlich S.D., Jamet E., Bolotin A., Renault P.;
RT   "Characterization of the two genes encoding histone-like proteins of
RT   the HU family in Lactococcus lactis IL1403.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE006285; AAK04599.1; -.
DR   EMBL; AF320916; AAK08221.1; -.
DR   PIR; E86687; E86687.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   218 AA;  24616 MW;  C38957B3E7A798CB CRC64;

  Query Match             48.2%;  Score 40;  DB 16;  Length 218;
  Best Local Similarity   70.0%;  Pred. No. 60;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          5 KQMTYKVYMS 14
              |:|||| |:|
Db        100 KEMTYKFYIS 109


RESULT 12
Q9U6Y8
ID   Q9U6Y8      PRELIMINARY;      PRT;  225 AA.
AC   Q9U6Y8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Fluorescent protein FP583.
OS   Discosoma sp.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
```

```
OC   Discosomatidae; Discosoma.
OX   NCBI_TaxID=86600;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99436614; PubMed=10504696;
RA   Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
RA   Markelov M.L., Lukyanov S.A.;
RT   "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL   Nat. Biotechnol. 17:969-973(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
RA   Markelov M.L., Lukyanov S.A.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF168419; AAF03369.1; -.
DR   PDB; 1G7K; 07-NOV-01.
DR   PDB; 1GGX; 06-DEC-00.
DR   GO; GO:0006091; P:energy pathways; IEA.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   225 AA;  25931 MW;  FBF9A5369778F689 CRC64;

  Query Match             48.2%;  Score 40;  DB 5;  Length 225;
  Best Local Similarity   52.9%;  Pred. No. 62;
  Matches    9;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              :|| : | :|| | |||
Db          6 NVIKEFMRFKVRMEGTV 22


RESULT 13
Q9GTJ7
ID   Q9GTJ7      PRELIMINARY;      PRT;   230 AA.
AC   Q9GTJ7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Red fluorescent protein.
GN   FP593.
OS   Discosoma sp. SSAL-2000.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
OX   NCBI_TaxID=137428;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20434599; PubMed=10981720;
RA   Fradkov A.F., Chen Y., Ding L., Barsova E.V., Matz M.V.,
RA   Lukyanov S.A.;
RT   "Novel fluorescent protein from Discosoma coral and its mutants
RT   possesses a unique far-red fluorescence.";
RL   FEBS Lett. 479:127-130(2000).
DR   EMBL; AF272711; AAG16224.1; -.
DR   HSSP; P42212; 1BFP.
DR   GO; GO:0006091; P:energy pathways; IEA.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   230 AA;  26370 MW;  5215B1B436D67E51 CRC64;

  Query Match             48.2%;  Score 40;  DB 5;  Length 230;
  Best Local Similarity   52.9%;  Pred. No. 63;
  Matches    9;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              :|| : | :|| | |||
Db          6 NVIKEFMRFKVRMEGTV 22


RESULT 14
Q97FT4
ID   Q97FT4      PRELIMINARY;      PRT;   263 AA.
AC   Q97FT4;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein CAC2643.
GN   CAC2643.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
DR   EMBL; AE007761; AAK80590.1; -.
DR   PIR; C97225; C97225.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   263 AA;  29220 MW;  41D648F7237A42AB CRC64;

  Query Match             48.2%;  Score 40;  DB 16;  Length 263;
  Best Local Similarity   43.8%;  Pred. No. 72;
  Matches    7;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              | : : :||| | ||:
Db        152 SGVVQDITYKAYTSGS 167


RESULT 15
Q7XDZ1
ID   Q7XDZ1      PRELIMINARY;      PRT;   268 AA.
AC   Q7XDZ1;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative AP2-domain DNA-binding protein.
GN   OSJNBA0094K20.6.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   The Rice Chromosome 10 Sequencing Consortium;
RT   "In-depth view of structure, activity, and evolution of rice
RT   chromosome 10.";
RL   Science 300:1566-1569(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE017099; AAP54000.1; -.
KW   DNA-binding.
SQ   SEQUENCE   268 AA;  30728 MW;  58DC0E0881BEA6E1 CRC64;

  Query Match             48.2%;  Score 40;  DB 10;  Length 268;
  Best Local Similarity   57.1%;  Pred. No. 74;
  Matches    8;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;
```

```
Qy          3 IAKQMTYKVYMSGT 16
              : | | ||||  ||
Db         94 LLKIMIYKVYADGT 107
```

Search completed: August 12, 2004, 06:19:35
Job time : 7.67418 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 6.44467 Seconds
(without alignments)
745.314 Million cell updates/sec

Title: US-09-890-463-2
Perfect score: 83
Sequence: 1 SVIAKQMTYKVYMSGTV 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	83	100.0	17	3	AAY97148	Aay97148 Pigment p
2	83	100.0	18	3	AAY97151	Aay97151 Pigment p
3	83	100.0	25	3	AAY97152	Aay97152 Pigment p
4	83	100.0	169	5	ABP69949	Abp69949 Colour Fa
5	83	100.0	169	5	ABP69944	Abp69944 Colour Fa
6	83	100.0	200	5	ABP69957	Abp69957 Colour Fa
7	83	100.0	220	5	ABP69941	Abp69941 Colour Fa
8	83	100.0	220	5	ABP69952	Abp69952 Colour Fa
9	83	100.0	220	5	ABP69925	Abp69925 Colour Fa
10	83	100.0	220	5	ABP69947	Abp69947 Colour Fa
11	83	100.0	220	5	ABP69959	Abp69959 Colour Fa
12	83	100.0	220	5	ABP69940	Abp69940 Colour Fa
13	83	100.0	220	5	ABP69943	Abp69943 Colour Fa
14	83	100.0	220	5	ABP69955	Abp69955 Colour Fa
15	83	100.0	220	5	ABP69929	Abp69929 Colour Fa
16	83	100.0	220	5	ABP69934	Abp69934 Colour Fa
17	83	100.0	220	5	ABP69958	Abp69958 Colour Fa
18	83	100.0	220	5	ABP69939	Abp69939 Colour Fa
19	83	100.0	220	5	ABP69953	Abp69953 Colour Fa
20	83	100.0	220	5	ABP69938	Abp69938 Colour Fa
21	83	100.0	220	5	ABP69945	Abp69945 Colour Fa
22	83	100.0	220	5	ABP69927	Abp69927 Colour Fa
23	83	100.0	220	5	ABP69946	Abp69946 Colour Fa
24	83	100.0	220	5	ABP69926	Abp69926 Colour Fa
25	83	100.0	220	5	ABP69956	Abp69956 Colour Fa
26	83	100.0	220	5	ABP69937	Abp69937 Colour Fa
27	83	100.0	220	5	ABP69932	Abp69932 Colour Fa
28	83	100.0	220	5	ABP69928	Abp69928 Colour Fa
29	83	100.0	220	5	ABP69931	Abp69931 Colour Fa
30	83	100.0	220	5	ABP69935	Abp69935 Colour Fa
31	83	100.0	220	5	ABP69936	Abp69936 Colour Fa
32	83	100.0	220	5	ABP69948	Abp69948 Colour Fa
33	83	100.0	220	5	ABP70037	Abp70037 Colour Fa
34	83	100.0	220	5	ABP69930	Abp69930 Colour Fa
35	83	100.0	222	5	ABP70028	Abp70028 Colour Fa
36	83	100.0	222	5	ABP70027	Abp70027 Colour Fa
37	83	100.0	223	5	ABP70033	Abp70033 Colour Fa
38	83	100.0	223	5	ABP70030	Abp70030 Colour Fa
39	83	100.0	223	5	ABP70029	Abp70029 Colour Fa
40	83	100.0	223	5	ABP70031	Abp70031 Colour Fa
41	83	100.0	223	5	ABP70032	Abp70032 Colour Fa
42	83	100.0	226	5	ABP70036	Abp70036 Colour Fa
43	83	100.0	231	3	AAY97149	Aay97149 Pigment p
44	83	100.0	231	5	ABP70025	Abp70025 Colour Fa
45	83	100.0	235	3	AAY97150	Aay97150 Pigment p

ALIGNMENTS

RESULT 1
AAY97148
ID AAY97148 standard; peptide; 17 AA.
XX
AC AAY97148;
XX
DT 04-DEC-2000 (first entry)
XX
DE Pigment protein from coral tissue N-terminal peptide 2.
XX
KW N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW UV filter.
XX
OS Acropora horrida.
XX
PN WO200046233-A1.
XX
PD 10-AUG-2000.
XX
PF 02-FEB-2000; 2000WO-AU000056.
XX
PR 02-FEB-1999; 99AU-00008463.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Hoegh-Guldberg O, Dove S;
XX
DR WPI; 2000-532892/48.
XX
PT Novel pigment protein derived from corals capable of emitting
PT fluorescence upon irradiation by incident light useful as tissue marker,
PT fluorescent marker or general dyestuff.
XX
PS Claim 4; Page 42; 49pp; English.
XX
CC The N-terminal peptides shown in AAY97147-48 are from pigment protein
CC from coral tissue (PPCT). PPCT is capable of emitting fluorescence upon
CC irradiation by incident light whose maximal absorbance is in the range of
CC 320-600 nm and a maximal fluorescence emission is in the range of 300-700
CC nm. PPCT may be used as a tissue marker, fluorescent marker (e.g. to
CC follow gene expression in transformed tissues) or general dyestuff (all
CC claimed). PPCT may also be used in sunscreen formulations or UV filters
CC (both claimed)
XX
SQ Sequence 17 AA;

Query Match 100.0%; Score 83; DB 3; Length 17;

```
  Best Local Similarity   100.0%;  Pred. No. 3.6e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTV 17


RESULT 2
AAY97151
ID   AAY97151 standard; peptide; 18 AA.
XX
AC   AAY97151;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Pigment protein from coral tissue N-terminal peptide 3.
XX
KW   N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW   tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW   UV filter.
XX
OS   Acropora aspera.
OS   Montipora caliculata.
OS   Porites murrayensis.
XX
PN   WO200046233-A1.
XX
PD   10-AUG-2000.
XX
PF   02-FEB-2000; 2000WO-AU000056.
XX
PR   02-FEB-1999;   99AU-00008463.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Hoegh-Guldberg O,  Dove S;
XX
DR   WPI; 2000-532892/48.
XX
PT   Novel pigment protein derived from corals capable of emitting
PT   fluorescence upon irradiation by incident light useful as tissue marker,
PT   fluorescent marker or general dyestuff.
XX
PS   Example 2; Page 18; 49pp; English.
XX
CC   The N-terminal peptides shown in AAY97151-52 are from pigment protein
CC   from coral tissue (PPCT). PPCT is capable of emitting fluorescence upon
CC   irradiation by incident light whose maximal absorbance is in the range of
CC   320-600 nm and a maximal fluorescence emission is in the range of 300-700
CC   nm. PPCT may be used as a tissue marker, fluorescent marker (e.g. to
CC   follow gene expression in transformed tissues) or general dyestuff (all
CC   claimed). PPCT may also be used in sunscreen formulations or UV filters
CC   (both claimed)
XX
SQ   Sequence 18 AA;

  Query Match             100.0%;  Score 83;  DB 3;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTV 17


RESULT 3
AAY97152
ID   AAY97152 standard; peptide; 25 AA.
XX
AC   AAY97152;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Pigment protein from coral tissue N-terminal peptide 4.
XX
KW   N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW   tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW   UV filter.
XX
OS   Porites lobata.
XX
PN   WO200046233-A1.
XX
PD   10-AUG-2000.
XX
PF   02-FEB-2000; 2000WO-AU000056.
XX
PR   02-FEB-1999;   99AU-00008463.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Hoegh-Guldberg O,  Dove S;
XX
DR   WPI; 2000-532892/48.
XX
PT   Novel pigment protein derived from corals capable of emitting
PT   fluorescence upon irradiation by incident light useful as tissue marker,
PT   fluorescent marker or general dyestuff.
XX
PS   Example 2; Page 18; 49pp; English.
XX
CC   The N-terminal peptides shown in AAY97151-52 are from pigment protein
CC   from coral tissue (PPCT). PPCT is capable of emitting fluorescence upon
CC   irradiation by incident light whose maximal absorbance is in the range of
CC   320-600 nm and a maximal fluorescence emission is in the range of 300-700
CC   nm. PPCT may be used as a tissue marker, fluorescent marker (e.g. to
CC   follow gene expression in transformed tissues) or general dyestuff (all
CC   claimed). PPCT may also be used in sunscreen formulations or UV filters
CC   (both claimed)
XX
SQ   Sequence 25 AA;

  Query Match             100.0%;  Score 83;  DB 3;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTV 17


RESULT 4
ABP69949
ID   ABP69949 standard; protein; 169 AA.
XX
AC   ABP69949;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 70.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Platygyra sp.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
```

PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 349; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 169 AA;

Query Match 100.0%; Score 83; DB 5; Length 169;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
|||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

RESULT 5
ABP69944
ID ABP69944 standard; protein; 169 AA.
XX
AC ABP69944;
XX
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 60.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Porites murrayensis.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 337; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 169 AA;

Query Match 100.0%; Score 83; DB 5; Length 169;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
|||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

RESULT 6
ABP69957
ID ABP69957 standard; protein; 200 AA.
XX
AC ABP69957;
XX
DT 06-AUG-2003 (revised)
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 84.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Montipora sp.
XX

PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 363-364; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 200 AA;

Query Match 100.0%; Score 83; DB 5; Length 200;
Best Local Similarity 100.0%; Pred. No. 6.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
|||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

RESULT 7
ABP69941
ID ABP69941 standard; protein; 220 AA.
XX
AC ABP69941;
XX
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 54.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Millepora sp.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 330-331; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 220 AA;

Query Match 100.0%; Score 83; DB 5; Length 220;
Best Local Similarity 100.0%; Pred. No. 7e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
|||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

RESULT 8
ABP69952
ID ABP69952 standard; protein; 220 AA.
XX
AC ABP69952;
XX
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 74.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;

KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Platygyra sp.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 351-352; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 220 AA;

Query Match 100.0%; Score 83; DB 5; Length 220;
Best Local Similarity 100.0%; Pred. No. 7e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

RESULT 9
ABP69925
ID ABP69925 standard; protein; 220 AA.
XX
AC ABP69925;
XX
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 22.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Acropora aspera.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 286-287; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 220 AA;

Query Match 100.0%; Score 83; DB 5; Length 220;
Best Local Similarity 100.0%; Pred. No. 7e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

RESULT 10
ABP69947
ID ABP69947 standard; protein; 220 AA.
XX
AC ABP69947;

XX
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 66.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Platygyra sp.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 344-345; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 220 AA;

Query Match 100.0%; Score 83; DB 5; Length 220;
Best Local Similarity 100.0%; Pred. No. 7e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
|||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

RESULT 11
ABP69959

ID ABP69959 standard; protein; 220 AA.
XX
AC ABP69959;
XX
DT 06-AUG-2003 (revised)
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 88.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Montipora sp.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 368-369; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 220 AA;

Query Match 100.0%; Score 83; DB 5; Length 220;
Best Local Similarity 100.0%; Pred. No. 7e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
|||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

```
RESULT 12
ABP69940
ID   ABP69940 standard; protein; 220 AA.
XX
AC   ABP69940;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 52.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Millepora sp.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 327-328; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences
XX
SQ   Sequence 220 AA;

  Query Match             100.0%;  Score 83;  DB 5;  Length 220;
  Best Local Similarity   100.0%;  Pred. No. 7e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTV 17


RESULT 13
ABP69943
ID   ABP69943 standard; protein; 220 AA.
XX
AC   ABP69943;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 58.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Porites murrayensis.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 335; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences
XX
SQ   Sequence 220 AA;

  Query Match             100.0%;  Score 83;  DB 5;  Length 220;
```

```
  Best Local Similarity   100.0%;  Pred. No. 7e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTV 17


RESULT 14
ABP69955
ID   ABP69955 standard; protein; 220 AA.
XX
AC   ABP69955;
XX
DT   06-AUG-2003  (revised)
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 80.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Pavona decussata.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 359; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 220 AA;

  Query Match           100.0%;  Score 83;  DB 5;  Length 220;
  Best Local Similarity   100.0%;  Pred. No. 7e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTV 17


RESULT 15
ABP69929
ID   ABP69929 standard; protein; 220 AA.
XX
AC   ABP69929;
XX
DT   06-AUG-2003  (revised)
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 30.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Acanthastrea sp.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 296-297; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
```

CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 220 AA;

Query Match 100.0%; Score 83; DB 5; Length 220;
Best Local Similarity 100.0%; Pred. No. 7e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

Search completed: August 12, 2004, 06:17:04
Job time : 6.44467 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:19:43 ; Search time 35.7766 Seconds
(without alignments)
149.169 Million cell updates/sec

Title: US-09-890-463-2
Perfect score: 83
Sequence: 1 SVIAKQMTYKVYMSGTV 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 1292805

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	51.8	107	10	US-09-994-595-43	Sequence 43, Appl
2	42	50.6	47	12	US-10-424-599-172498	Sequence 172498,
3	41	49.4	225	14	US-10-315-920-6	Sequence 6, Appli
4	41	49.4	225	15	US-10-442-148A-7	Sequence 7, Appli
5	41	49.4	239	15	US-10-442-148A-8	Sequence 8, Appli
6	40	48.2	26	14	US-10-081-864-25	Sequence 25, Appl
7	40	48.2	205	13	US-10-006-922-46	Sequence 46, Appl
8	40	48.2	225	9	US-09-999-745-67	Sequence 67, Appl
9	40	48.2	225	10	US-09-866-538-12	Sequence 12, Appl
10	40	48.2	225	10	US-09-794-308-12	Sequence 12, Appl
11	40	48.2	225	10	US-09-865-291-12	Sequence 12, Appl
12	40	48.2	225	12	US-10-132-067-4	Sequence 4, Appli
13	40	48.2	225	13	US-10-006-922-12	Sequence 12, Appl
14	40	48.2	225	13	US-10-006-922-44	Sequence 44, Appl
15	40	48.2	225	14	US-10-081-864-8	Sequence 8, Appli
16	40	48.2	225	14	US-10-081-864-12	Sequence 12, Appl
17	40	48.2	225	14	US-10-121-258-1	Sequence 1, Appli
18	40	48.2	225	14	US-10-121-258-20	Sequence 20, Appl
19	40	48.2	225	14	US-10-315-920-2	Sequence 2, Appli
20	40	48.2	225	14	US-10-315-920-4	Sequence 4, Appli
21	40	48.2	225	15	US-10-370-570-56	Sequence 56, Appl
22	40	48.2	225	15	US-10-406-618-32	Sequence 32, Appl
23	40	48.2	225	16	US-10-433-640-13	Sequence 13, Appl
24	40	48.2	226	16	US-10-724-178-12	Sequence 12, Appl
25	40	48.2	230	13	US-10-006-922-18	Sequence 18, Appl
26	40	48.2	230	14	US-10-161-403-40	Sequence 40, Appl
27	40	48.2	236	16	US-10-314-936-2	Sequence 2, Appli
28	40	48.2	236	16	US-10-314-936-4	Sequence 4, Appli
29	40	48.2	240	14	US-10-152-296-2	Sequence 2, Appli
30	40	48.2	240	16	US-10-739-656-2	Sequence 2, Appli
31	40	48.2	326	16	US-10-437-963-112625	Sequence 112625,
32	40	48.2	487	15	US-10-343-977-1	Sequence 1, Appli
33	40	48.2	506	15	US-10-343-977-2	Sequence 2, Appli
34	40	48.2	545	14	US-10-214-932-52	Sequence 52, Appl
35	40	48.2	547	15	US-10-343-977-3	Sequence 3, Appli
36	40	48.2	548	14	US-10-214-932-76	Sequence 76, Appl
37	39	47.0	107	10	US-09-994-595-80	Sequence 80, Appl
38	39	47.0	133	12	US-10-424-599-224455	Sequence 224455,
39	39	47.0	213	12	US-10-425-114-67457	Sequence 67457, A
40	39	47.0	215	14	US-10-155-809-8	Sequence 8, Appli
41	39	47.0	222	15	US-10-406-618-17	Sequence 17, Appl
42	39	47.0	225	16	US-10-423-688A-41	Sequence 41, Appl
43	39	47.0	226	9	US-09-976-673-8	Sequence 8, Appli
44	39	47.0	226	9	US-09-976-673-10	Sequence 10, Appl
45	39	47.0	226	9	US-09-976-673-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-09-994-595-43
; Sequence 43, Application US/09994595
; Publication No. US20030039981A1
; GENERAL INFORMATION:
; APPLICANT: Bhattacharjee, J.
; APPLICANT: Suvarna, Kalavati
; APPLICANT: Bhattacherjee, Vasker
; TITLE OF INVENTION: METHODS AND REAGENTS FOR DETECTING FUNGAL PATHOGENS IN
; TITLE OF INVENTION: A BIOLOGICAL SAMPLE
; FILE REFERENCE: 96,247-A
; CURRENT APPLICATION NUMBER: US/09/994,595
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 08/650,809
; PRIOR FILING DATE: 1997-05-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Microsoft Word 97
; SEQ ID NO 43
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Polypeptide segment of ACVS_EMENI shown in Figure 4.
US-09-994-595-43

Query Match 51.8%; Score 43; DB 10; Length 107;
Best Local Similarity 50.0%; Pred. No. 4.6;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 SVIAKQMTYKVYMSGT 16
              |: :||: |  | |||
Db          1 SLTSKQLAYVTYTSGT 16
```

RESULT 2
US-10-424-599-172498
; Sequence 172498, Application US/10424599
; Publication No. US20040031072A1

```
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 172498
;   LENGTH: 47
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(47)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_126782C.1.pep
US-10-424-599-172498

  Query Match             50.6%;  Score 42;  DB 12;  Length 47;
  Best Local Similarity   50.0%;  Pred. No. 2.8;
  Matches    7;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 IAKQMTYKVYMSGT 16
              :|:|  | ::|| ||
Db         16 VARQPTIRIYMLGT 29


RESULT 3
US-10-315-920-6
; Sequence 6, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
;  APPLICANT: Fradkov, Arcady Fedorovich
;  APPLICANT:  Terskikh, Alexey
;  TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
;  TITLE OF INVENTION:  FOR THEIR USE
;  FILE REFERENCE: CLON-077CIP
;  CURRENT APPLICATION NUMBER: US/10/315,920
;  CURRENT FILING DATE:  2002-12-09
;  PRIOR APPLICATION NUMBER: 60/211,607
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: PCT/US01/19097
;  PRIOR FILING DATE: 2001-06-13
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 225
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: variant of sequence from Discosoma sp.
US-10-315-920-6

  Query Match             49.4%;  Score 41;  DB 14;  Length 225;
  Best Local Similarity   52.9%;  Pred. No. 25;
  Matches    9;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              :|| : | :|| | |||
Db          6 NVITEFMRFKVRMEGTV 22


RESULT 4
US-10-442-148A-7
; Sequence 7, Application US/10442148A
; Publication No. US20040014242A1
; GENERAL INFORMATION:
;  APPLICANT: IWAKURA, MASAHIRO
;  APPLICANT:  HIROTA, KIYONORI
;  TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ORIENTATION-CONTROLLED PROTEIN AND
;  TITLE OF INVENTION:  PROCESS FOR ARRAYING AND IMMOBILIZING PROTEIN USING THE SAME
;  FILE REFERENCE: 04583.0103-00000
;  CURRENT APPLICATION NUMBER: US/10/442,148A
;  CURRENT FILING DATE:  2003-05-21
;  PRIOR APPLICATION NUMBER: JP 2002-148950
;  PRIOR FILING DATE: 2002-05-23
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
;   LENGTH: 225
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic protein sequence
US-10-442-148A-7

  Query Match             49.4%;  Score 41;  DB 15;  Length 225;
  Best Local Similarity   52.9%;  Pred. No. 25;
  Matches    9;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              :|| : | :|| | |||
Db          6 NVITEFMRFKVRMEGTV 22


RESULT 5
US-10-442-148A-8
; Sequence 8, Application US/10442148A
; Publication No. US20040014242A1
; GENERAL INFORMATION:
;  APPLICANT: IWAKURA, MASAHIRO
;  APPLICANT:  HIROTA, KIYONORI
;  TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ORIENTATION-CONTROLLED PROTEIN AND
;  TITLE OF INVENTION:  PROCESS FOR ARRAYING AND IMMOBILIZING PROTEIN USING THE SAME
;  FILE REFERENCE: 04583.0103-00000
;  CURRENT APPLICATION NUMBER: US/10/442,148A
;  CURRENT FILING DATE:  2003-05-21
;  PRIOR APPLICATION NUMBER: JP 2002-148950
;  PRIOR FILING DATE: 2002-05-23
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
;   LENGTH: 239
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic protein sequence
US-10-442-148A-8

  Query Match             49.4%;  Score 41;  DB 15;  Length 239;
  Best Local Similarity   52.9%;  Pred. No. 27;
  Matches    9;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              :|| : | :|| | |||
Db          6 NVITEFMRFKVRMEGTV 22


RESULT 6
US-10-081-864-25
; Sequence 25, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
;  APPLICANT: Lukyanov, Sergey
;  APPLICANT:  Lukyanov, Konstantin
;  APPLICANT:  Yanushevich, Yuriy
;  APPLICANT:  Savistky, Alexandr
;  APPLICANT:  Fradkov, Arcady
;  TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
;  TITLE OF INVENTION:  Methods for Using the Same
;  FILE REFERENCE: CLON-067
```

; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: non-aggregating mutant fragment
US-10-081-864-25

Query Match 48.2%; Score 40; DB 14; Length 26;
Best Local Similarity 52.9%; Pred. No. 3.3;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
:|| : | :|| | |||
Db 6 NVIKEFMRFKVRMEGTV 22

RESULT 7
US-10-006-922-46
; Sequence 46, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Terskikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Discosoma species
US-10-006-922-46

Query Match 48.2%; Score 40; DB 13; Length 205;
Best Local Similarity 52.9%; Pred. No. 34;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
:|| : | :|| | |||
Db 6 NVIKEFMRFKVRMEGTV 22

RESULT 8
US-09-999-745-67
; Sequence 67, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 67
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-999-745-67

Query Match 48.2%; Score 40; DB 9; Length 225;
Best Local Similarity 52.9%; Pred. No. 38;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
:|| : | :|| | |||
Db 6 NVIKEFMRFKVRMEGTV 22

RESULT 9
US-09-866-538-12
; Sequence 12, Application US/09866538
; Publication No. US20030032088A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-866-538-12

Query Match 48.2%; Score 40; DB 10; Length 225;
Best Local Similarity 52.9%; Pred. No. 38;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
:|| : | :|| | |||
Db 6 NVIKEFMRFKVRMEGTV 22

RESULT 10
US-09-794-308-12
; Sequence 12, Application US/09794308
; Publication No. US20030170911A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: ZACHARIAS, David
; APPLICANT: BAIRD, Geoffrey
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530
; CURRENT APPLICATION NUMBER: US/09/794,308
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 225

```
;   TYPE: PRT
;   ORGANISM: Discosoma sp.
US-09-794-308-12

  Query Match             48.2%;  Score 40;  DB 10;  Length 225;
  Best Local Similarity   52.9%;  Pred. No. 38;
  Matches    9;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              :|| : | :|| | |||
Db          6 NVIKEFMRFKVRMEGTV 22


RESULT 11
US-09-865-291-12
; Sequence 12, Application US/09865291
; Publication No. US20030186229A1
; GENERAL INFORMATION:
;  APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
;  APPLICANT:  TSIEN, Roger
;  APPLICANT:  TING, Alice
;  APPLICANT:  ZHANG, Jin
;  TITLE OF INVENTION: EMISSION RATIOMETRIC INDICATORS OF PHOSPHORYLATION
;  FILE REFERENCE: REGEN1550
;  CURRENT APPLICATION NUMBER: US/09/865,291
;  CURRENT FILING DATE:  2001-05-24
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
;   LENGTH: 225
;   TYPE: PRT
;   ORGANISM: Discosoma sp.
US-09-865-291-12

  Query Match             48.2%;  Score 40;  DB 10;  Length 225;
  Best Local Similarity   52.9%;  Pred. No. 38;
  Matches    9;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              :|| : | :|| | |||
Db          6 NVIKEFMRFKVRMEGTV 22


RESULT 12
US-10-132-067-4
; Sequence 4, Application US/10132067
; Publication No. US20030203355A1
; GENERAL INFORMATION:
;  APPLICANT: Bradbury, Andrew
;  APPLICANT:  Zeytun, Ahmet
;  APPLICANT:  Waldo, Geoffrey
;  APPLICANT:  The Regents of the University of California
;  TITLE OF INVENTION: Fluorobodies:  Binding Ligands With Intrinsic
;  TITLE OF INVENTION:  Fluorescence
;  FILE REFERENCE: 021362-000600US
;  CURRENT APPLICATION NUMBER: US/10/132,067
;  CURRENT FILING DATE:  2002-04-24
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 225
;   TYPE: PRT
;   ORGANISM: Discosoma sp.
;   FEATURE:
;   OTHER INFORMATION: red fluorescent protein (dsRED)
US-10-132-067-4

  Query Match             48.2%;  Score 40;  DB 12;  Length 225;
  Best Local Similarity   52.9%;  Pred. No. 38;
  Matches    9;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              :|| : | :|| | |||
Db          6 NVIKEFMRFKVRMEGTV 22


RESULT 13
US-10-006-922-12
; Sequence 12, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
;  APPLICANT: Lukyanov, Sergey A
;  APPLICANT:  Fradkov, Arcady F.
;  APPLICANT:  Labas, Yulii A.
;  APPLICANT:  Matz, Mikhail V.
;  APPLICANT:  Terskikh, Alexey
;  TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
;  TITLE OF INVENTION:  Methods for Using the Same
;  FILE REFERENCE: CLON-035CIP
;  CURRENT APPLICATION NUMBER: US/10/006,922
;  CURRENT FILING DATE:  2001-12-04
;  PRIOR APPLICATION NUMBER: 09/120,330
;  PRIOR FILING DATE: 1998-12-11
;  PRIOR APPLICATION NUMBER: 09/457,898
;  PRIOR FILING DATE: 1999-12-09
;  PRIOR APPLICATION NUMBER: 09/458,144
;  PRIOR FILING DATE: 1999-12-09
;  PRIOR APPLICATION NUMBER: 09/458,477
;  PRIOR FILING DATE: 1999-12-09
;  PRIOR APPLICATION NUMBER: 09/457,556
;  PRIOR FILING DATE: 1999-12-09
;  PRIOR APPLICATION NUMBER: 09/444,338
;  PRIOR FILING DATE: 1999-11-19
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 225
;   TYPE: PRT
;   ORGANISM: Discosoma species
US-10-006-922-12

  Query Match             48.2%;  Score 40;  DB 13;  Length 225;
  Best Local Similarity   52.9%;  Pred. No. 38;
  Matches    9;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              :|| : | :|| | |||
Db          6 NVIKEFMRFKVRMEGTV 22


RESULT 14
US-10-006-922-44
; Sequence 44, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
;  APPLICANT: Lukyanov, Sergey A
;  APPLICANT:  Fradkov, Arcady F.
;  APPLICANT:  Labas, Yulii A.
;  APPLICANT:  Matz, Mikhail V.
;  APPLICANT:  Terskikh, Alexey
;  TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
;  TITLE OF INVENTION:  Methods for Using the Same
;  FILE REFERENCE: CLON-035CIP
;  CURRENT APPLICATION NUMBER: US/10/006,922
;  CURRENT FILING DATE:  2001-12-04
;  PRIOR APPLICATION NUMBER: 09/120,330
;  PRIOR FILING DATE: 1998-12-11
;  PRIOR APPLICATION NUMBER: 09/457,898
;  PRIOR FILING DATE: 1999-12-09
;  PRIOR APPLICATION NUMBER: 09/458,144
;  PRIOR FILING DATE: 1999-12-09
;  PRIOR APPLICATION NUMBER: 09/458,477
;  PRIOR FILING DATE: 1999-12-09
;  PRIOR APPLICATION NUMBER: 09/457,556
;  PRIOR FILING DATE: 1999-12-09
;  PRIOR APPLICATION NUMBER: 09/444,338
```

; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: hybrid construct
US-10-006-922-44

Query Match 48.2%; Score 40; DB 13; Length 225;
Best Local Similarity 52.9%; Pred. No. 38;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 SVIAKQMTYKVYMSGTV 17
              :|| : | :|| | |||
Db          6 NVIKEFMRFKVRMEGTV 22
```

RESULT 15
US-10-081-864-8
; Sequence 8, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savistky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-081-864-8

Query Match 48.2%; Score 40; DB 14; Length 225;
Best Local Similarity 52.9%; Pred. No. 38;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 SVIAKQMTYKVYMSGTV 17
              :|| : | :|| | |||
Db          6 NVIKEFMRFKVRMEGTV 22
```

Search completed: August 12, 2004, 06:51:20
Job time : 35.7766 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:34:08 ; Search time 16 Seconds
(without alignments)
30.060 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 206

Minimum DB seq length: 0
Maximum DB seq length: 5

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	10	47.6	4	2	PL0146	carbon-monoxide de
2	9	42.9	4	2	I40505	hypothetical prote
3	9	42.9	4	2	A27897	glucan 1,4-alpha-g
4	9	42.9	5	2	S11127	phosphoprotein, bo
5	8	38.1	4	2	S17255	ribosomal protein
6	8	38.1	4	2	T30569	hypothetical prote
7	8	38.1	4	2	E44823	synaptosomal-assoc
8	8	38.1	5	2	A60521	glycogen phosphory
9	8	38.1	5	2	T14908	hypothetical prote
10	7	33.3	4	2	I61883	protamine P1 - ora
11	7	33.3	4	2	PT0551	T-cell receptor be
12	7	33.3	4	2	PT0697	T-cell receptor be
13	7	33.3	4	2	I37013	protamine P1 - Cer
14	7	33.3	4	2	I84439	protamine P1 - sav
15	7	33.3	5	2	I39964	ribosomal protein
16	7	33.3	5	2	I39966	ribosomal protein
17	7	33.3	5	2	I39965	ribosomal protein
18	7	33.3	5	2	T14910	hypothetical prote
19	7	33.3	5	2	S55237	zinc-binding prote
20	7	33.3	5	2	PT0597	T-cell receptor be
21	6	28.6	4	2	S18401	thyroglobulin - do
22	6	28.6	4	2	A61300	22K superhelical D
23	6	28.6	4	2	PT0534	T-cell receptor be
24	6	28.6	4	2	PT0721	T-cell receptor be
25	6	28.6	5	2	A44955	alkanal monooxygen
26	6	28.6	5	2	B61168	cocoonase (EC 3.4.
27	6	28.6	5	2	I50385	myosin light chain
28	6	28.6	5	2	PT0278	Ig heavy chain CRD
29	6	28.6	5	2	PT0295	Ig heavy chain CRD
30	6	28.6	5	2	S62883	seminal plasma pro
31	6	28.6	5	2	B44817	34.5K structural p
32	6	28.6	5	2	D44817	35K structural pro
33	5	23.8	3	3	GKHU	growth-modulating
34	5	23.8	3	3	A60898	bursin - chicken
35	5	23.8	3	3	S13894	histidinol dehydro
36	5	23.8	3	3	E37196	bradykinin-potenti
37	5	23.8	3	3	F37196	bradykinin-potenti
38	5	23.8	3	3	PT0578	T-cell receptor be
39	5	23.8	4	2	A02147	phagocytosis-stimu
40	5	23.8	4	2	A37832	phenol 2-monooxyge
41	5	23.8	4	2	I40870	phospholipase C (E
42	5	23.8	4	2	I40804	endoglucanase F -
43	5	23.8	4	2	T46627	hypothetical prote
44	5	23.8	4	2	JQ1273	neuropeptide Antho
45	5	23.8	4	2	S39390	myosin-light-chain

ALIGNMENTS

RESULT 1
PL0146
carbon-monoxide dehydrogenase (EC 1.2.99.2) small chain - Pseudomonas carboxydohydrogena
C;Species: Pseudomonas carboxydohydrogena
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 28-Apr-1993
C;Accession: PL0146
R;Kraut, M.; Hugendieck, I.; Herwig, S.; Meyer, O.
Arch. Microbiol. 152, 335-341, 1989
A;Title: Homology and distribution of CO dehydrogenase structural genes in carboxydotroph
A;Reference number: PL0138; MUID:90055678; PMID:2818128
A;Accession: PL0146
A;Molecule type: protein
A;Residues: 1-4 <KRA>
C;Comment: Carbon-monoxide dehydrogenase consists of three polypeptide chains: large, med
C;Keywords: oxidoreductase

Query Match 47.6%; Score 10; DB 2; Length 4;
Best Local Similarity 66.7%; Pred. No. 2.8e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          3 IAK 5
              :||
Db          1 MAK 3
```

RESULT 2
I40505
hypothetical protein 3 (4 aa) - Bacillus stearothermophilus
C;Species: Bacillus stearothermophilus
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999
C;Accession: I40505
R;Waye, M.M.; Winter, G.
Eur. J. Biochem. 158, 505-510, 1986
A;Title: A transcription terminator in the 5' non-coding region of the tyrosyl tRNA synth
A;Reference number: I40503; MUID:86274732; PMID:3525162
A;Accession: I40505
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4 <RES>
A;Cross-references: EMBL:X04193; NID:g40233; PIDN:CAA27783.1; PID:g580944

Query Match 42.9%; Score 9; DB 2; Length 4;
Best Local Similarity 25.0%; Pred. No. 2.8e+05;
Matches 1; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy          2 VIAK 5
              :::|
Db          1 MLSK 4
```

RESULT 3
A27897

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - Aspergillus phoenicis (fragment)
N;Alternate names: glucoamylase
C;Species: Aspergillus phoenicis
C;Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 06-Dec-1996
C;Accession: A27897
R;Inokuchi, N.; Takahashi, T.; Irie, M.
J. Biochem. 90, 1055-1067, 1981
A;Title: Purification and characterization of a minor glucoamylase from Aspergillus sait
A;Reference number: A27897; MUID:82075730; PMID:6796572
A;Note: Aspergillus saitoi
A;Accession: A27897
A;Molecule type: protein
A;Residues: 1-4 <INO>
C;Keywords: glycosidase; hydrolase; polysaccharide degradation

  Query Match 42.9%; Score 9; DB 2; Length 4;
  Best Local Similarity 66.7%; Pred. No. 2.8e+05;
  Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVI 3
:||
Db 1 AVI 3

RESULT 4
S11127
phosphoprotein, bone - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 04-Mar-2000
C;Accession: S11127; S11128
R;Mikuni-Takagaki, Y.; Glimcher, M.J.
Biochem. J. 268, 585-591, 1990
A;Title: Post-translational processing of chicken bone phosphoproteins. Identification c
A;Reference number: S11127; MUID:90303246; PMID:2363696
A;Accession: S11127
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-5 <MIK1>
A;Accession: S11128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 'X',2-5 <MIK2>
C;Keywords: phosphoprotein

  Query Match 42.9%; Score 9; DB 2; Length 5;
  Best Local Similarity 33.3%; Pred. No. 2.8e+05;
  Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IAK 5
::|
Db 3 VSK 5

RESULT 5
S17255
ribosomal protein YmL1, mitochondrial, questionable - yeast (Saccharomyces cerevisiae) (
C;Species: Saccharomyces cerevisiae
A;Variety: strain 07173
C;Date: 23-Apr-1993 #sequence_revision 14-Sep-1994 #text_change 09-May-1997
C;Accession: S17255
R;Grohmann, L.; Graack, H.R.; Kruft, V.; Choli, T.; Goldschmidt-Reisin, S.; Kitakawa, M.
FEBS Lett. 284, 51-56, 1991
A;Title: Extended N-terminal sequencing of proteins of the large ribosomal subunit from
A;Reference number: S17255; MUID:91285106; PMID:2060626
A;Accession: S17255
A;Molecule type: protein
A;Residues: 1-4 <GRO>
C;Comment: A coding region for this protein could not be identified in the genome of Sac
C;Genetics:
A;Genome: nuclear
C;Keywords: mitochondrion; protein biosynthesis; ribosome

  Query Match 38.1%; Score 8; DB 2; Length 4;
  Best Local Similarity 100.0%; Pred. No. 2.8e+05;
  Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SV 2
||
Db 1 SV 2

RESULT 6
T30569
hypothetical protein - Emericella nidulans
C;Species: Emericella nidulans, Aspergillus nidulans
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C;Accession: T30569
R;Morrice, J.; MacKenzie, D.A.; Parr, A.J.; Archer, D.B.
Curr. Genet. 34, 379-385, 1998
A;Title: Isolation and characterisation of the acetyl-CoA carboxylase gene from Aspergil
A;Reference number: Z20869; MUID:99087906; PMID:9871120
A;Accession: T30569
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4 <MOR>
A;Cross-references: EMBL:Y15996; NID:e1285512; PID:e1218041; PIDN:CAA75927.1

  Query Match 38.1%; Score 8; DB 2; Length 4;
  Best Local Similarity 100.0%; Pred. No. 2.8e+05;
  Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SV 2
||
Db 3 SV 4

RESULT 7
E44823
synaptosomal-associated protein SNAP-25 peptide 1 - rabbit (fragment)
N;Alternate names: superprotein peptide 1
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 31-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 15-Jun-1996
C;Accession: E44823
R;Loewy, A.; Liu, W.S.; Baitinger, C.; Willard, M.B.
J. Neurosci. 11, 3412-3421, 1991
A;Title: The major 35S-methionine-labeled rapidly transported protein (superprotein) is
A;Reference number: A44823; MUID:92044785; PMID:1941090
A;Accession: E44823
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-4 <LOE>
A;Experimental source: visual tissue
A;Note: sequence extracted from NCBI backbone (NCBIP:64247)
C;Keywords: membrane trafficking

  Query Match 38.1%; Score 8; DB 2; Length 4;
  Best Local Similarity 25.0%; Pred. No. 2.8e+05;
  Matches 1; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VIAK 5
:: |
Db 1 IMEK 4

RESULT 8
A60521
glycogen phosphorylase (EC 2.4.1.1), muscle - mullet (Liza ramada) (fragment)
N;Alternate names: glycogen phosphorylase b
C;Species: Liza ramada
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 18-Aug-2003
C;Accession: A60521
R;Bonamusa, L.; Baanante, I.V.
Comp. Biochem. Physiol. B 95, 295-301, 1990
A;Title: Purification and characterization of glycogen phosphorylase B from skeletal mus
A;Reference number: A60521; MUID:90227907; PMID:2109669
A;Accession: A60521

A;Molecule type: protein
A;Residues: 1-5 <BON>
C;Superfamily: glucan phosphorylase
C;Keywords: glycosyltransferase; hexosyltransferase; phosphoprotein
F;3/Binding site: phosphate (Ser) (covalent) (by phosphorylase b kinase) #status experim

Query Match 38.1%; Score 8; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SV 2
||
Db 3 SV 4

RESULT 9
T14908
hypothetical protein - parsley
C;Species: Petroselinum crispum (parsley)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T14908
R;Kircher, S.; Ledger, S.; Hayashi, H.; Weisshaar, B.; Schafer, E.; Frohnmeyer, H.
Mol. Gen. Genet. 257, 595-605, 1998
A;Title: CPRF4a, a novel plant bZIP protein of the CPRF family: comparative analysis of
A;Reference number: Z18261; MUID:98265918; PMID:9604882
A;Accession: T14908
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-5 <KIR>
A;Cross-references: EMBL:Y10809; NID:g3336901; PIDN:CAA71767.1; PID:g3336902
A;Experimental source: Hamburger Schnitt

Query Match 38.1%; Score 8; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SV 2
||
Db 3 SV 4

RESULT 10
I61883
protamine P1 - orangutan (fragment)
C;Species: Pongo pygmaeus (orangutan)
C;Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 21-Jul-2000
C;Accession: I61883
R;Queralt, R.; Oliva, R.
Gene 133, 197-204, 1993
A;Title: Identification of conserved potential regulatory sequences of the protamine-enc
A;Reference number: I37013; MUID:94040810; PMID:8224908
A;Accession: I61883
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4 <RES>
A;Cross-references: EMBL:Z12146; NID:g38156; PIDN:CAA78130.1; PID:g4379372

Query Match 33.3%; Score 7; DB 2; Length 4;
Best Local Similarity 33.3%; Pred. No. 2.8e+05;
Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IAK 5
:|:
Db 1 MAR 3

RESULT 11
PT0551
T-cell receptor beta chain V-D-J region (126-1CG) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0551
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601; PMID:1711558
A;Accession: PT0551
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-4 <FEE>
A;Experimental source: day 18 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor

Query Match 33.3%; Score 7; DB 2; Length 4;
Best Local Similarity 50.0%; Pred. No. 2.8e+05;
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SV 2
|:
Db 3 SI 4

RESULT 12
PT0697
T-cell receptor beta chain V-D-J region (135-1BF) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0697
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601; PMID:1711558
A;Accession: PT0697
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-4 <FEE>
A;Experimental source: newborn thymus, strain BALB/c
C;Keywords: T-cell receptor

Query Match 33.3%; Score 7; DB 2; Length 4;
Best Local Similarity 50.0%; Pred. No. 2.8e+05;
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SV 2
|:
Db 2 SI 3

RESULT 13
I37013
protamine P1 - Cercopithecus patas (fragment)
C;Species: Cercopithecus patas
C;Date: 19-Mar-1997 #sequence_revision 07-Nov-1997 #text_change 21-Jul-2000
C;Accession: I37013
R;Queralt, R.; Oliva, R.
Gene 133, 197-204, 1993
A;Title: Identification of conserved potential regulatory sequences of the protamine-enc
A;Reference number: I37013; MUID:94040810; PMID:8224908
A;Accession: I37013
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4 <RES>
A;Cross-references: EMBL:Z12150; NID:g22814; PIDN:CAA78134.1; PID:g4377415

Query Match 33.3%; Score 7; DB 2; Length 4;
Best Local Similarity 33.3%; Pred. No. 2.8e+05;
Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IAK 5
:|:
Db 1 MAR 3

RESULT 14
I84439
protamine P1 - savannah baboon (fragment)

```
C;Species: Papio hamadryas doguera (savannah baboon)
C;Date: 19-Mar-1997 #sequence_revision 07-Nov-1997 #text_change 21-Jul-2000
C;Accession: I84439
R;Queralt, R.; Oliva, R.
Gene 133, 197-204, 1993
A;Title: Identification of conserved potential regulatory sequences of the protamine-enc
A;Reference number: I37013; MUID:94040810; PMID:8224908
A;Accession: I84439
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4 <RES>
A;Cross-references: EMBL:Z12147; NID:g38134; PIDN:CAA78131.1; PID:g4379349

  Query Match             33.3%;  Score 7;  DB 2;  Length 4;
  Best Local Similarity   33.3%;  Pred. No. 2.8e+05;
  Matches    1;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 IAK 5
              :|:
Db          1 MAR 3


RESULT 15
I39964
ribosomal protein S4 - Bacillus circulans (fragment)
C;Species: Bacillus circulans
C;Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 19-Jul-1996
C;Accession: I39964
R;Grundy, F.J.; Henkin, T.M.
J. Bacteriol. 174, 6763-6770, 1992
A;Title: Characterization of the Bacillus subtilis rpsD regulatory target site.
A;Reference number: I39963; MUID:93015735; PMID:1400226
A;Accession: I39964
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5 <RES>
A;Cross-references: GB:M99041; NID:g143471
C;Genetics:
A;Gene: rpsD

  Query Match             33.3%;  Score 7;  DB 2;  Length 5;
  Best Local Similarity   33.3%;  Pred. No. 2.8e+05;
  Matches    1;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 IAK 5
              :|:
Db          1 MAR 3


Search completed: August 12, 2004, 06:55:20
Job time : 17 secs
```

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:21:13 ; Search time 13 Seconds
(without alignments)
20.027 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 38

Minimum DB seq length: 0
Maximum DB seq length: 5

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	10	47.6	4	1	DCMS_PSECH	P19918 pseudomonas
2	8	38.1	4	1	RM01_YEAST	P36515 saccharomyc
3	6	28.6	5	1	UF01_MOUSE	P38639 mus musculu
4	5	23.8	3	1	GRWM_HUMAN	P01157 homo sapien
5	5	23.8	3	1	LUXE_VIBFI	P24272 vibrio fisc
6	5	23.8	4	1	FFKA_ANTEL	P58705 anthopleura
7	5	23.8	4	1	TUFT_HUMAN	P01858 homo sapien
8	5	23.8	5	1	BIOB_CITFR	P12997 citrobacter
9	5	23.8	5	1	BPP7_BOTIN	P30425 bothrops in
10	5	23.8	5	1	EI04_LITRU	P82100 litoria rub
11	5	23.8	5	1	TRM3_ECOLI	P13973 escherichia
12	4	19.0	4	1	ACH1_ACHFU	P35904 achatina fu
13	4	19.0	4	1	EOSI_HUMAN	P02731 homo sapien
14	4	19.0	4	1	FYRI_ANTEL	P58706 anthopleura
15	4	19.0	4	1	OCP3_OCTMI	P58649 octopus min
16	4	19.0	5	1	AL14_CARMA	P81817 carcinus ma
17	4	19.0	5	1	EI03_LITRU	P82099 litoria rub
18	4	19.0	5	1	FARP_ARTTR	P41853 artioposthi
19	4	19.0	5	1	PSK_DAUCA	P58261 daucus caro
20	4	19.0	5	1	RE11_LITRU	P82070 litoria rub
21	4	19.0	5	1	RE21_LITRU	P82071 litoria rub
22	4	19.0	5	1	RE31_LITRU	P82072 litoria rub
23	4	19.0	5	1	RE32_LITRU	P82073 litoria rub
24	4	19.0	5	1	SUGA_ACHDO	P19991 acheta dome
25	4	19.0	5	1	TPIS_CANFA	P54714 canis famil
26	4	19.0	5	1	UC22_MAIZE	P80628 zea mays (m
27	4	19.0	5	1	UXA4_CHLTR	P38005 chlamydia t
28	2	9.5	4	1	FAR3_HIRME	P42562 hirudo medi
29	2	9.5	4	1	FAR4_HIRME	P42563 hirudo medi
30	2	9.5	4	1	FLRF_HIRME	P42561 hirudo medi
31	2	9.5	4	1	FLRN_ANTEL	P58707 anthopleura
32	2	9.5	4	1	FMRF_MACNI	P01162 macrocallis
33	2	9.5	5	1	PRCT_PERAM	P01373 periplaneta
34	1	4.8	3	1	THYL_PIG	P01151 sus scrofa
35	1	4.8	4	1	DCML_PSECH	P19916 pseudomonas
36	1	4.8	5	1	BIOA_CITFR	P13071 citrobacter
37	0	0.0	4	1	OCP1_OCTMI	P58648 octopus min
38	0	0.0	5	1	PAP2_PARMA	P81864 pardachirus

ALIGNMENTS

RESULT 1
DCMS_PSECH
ID DCMS_PSECH STANDARD; PRT; 4 AA.
AC P19918;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbon monoxide dehydrogenase small chain (EC 1.2.99.2) (CO
DE dehydrogenase subunit S) (CO-DH S) (Fragment).
GN CUTS.
OS Pseudomonas carboxydohydrogena.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae.
OX NCBI_TaxID=290;
RN [1]
RP SEQUENCE.
RX MEDLINE=90055678; PubMed=2818128;
RA Kraut M., Hugendieck I., Herwig S., Meyer O.;
RT "Homology and distribution of CO dehydrogenase structural genes in
RT carboxydotrophic bacteria.";
RL Arch. Microbiol. 152:335-341(1989).
CC -!- FUNCTION: Catalyzes the oxidation of carbon monoxide to carbon
CC dioxide.
CC -!- CATALYTIC ACTIVITY: CO + H(2)O + acceptor = CO(2) + reduced
CC acceptor.
CC -!- COFACTOR: Binds 2 2Fe-2S clusters.
CC -!- SUBUNIT: CONSISTS OF THREE POLYPEPTIDE CHAINS: LARGE, MEDIUM, AND
CC SMALL.
DR PIR; PL0146; PL0146.
KW Oxidoreductase; Metal-binding; Iron-sulfur; Iron; 2Fe-2S.
FT NON_TER 4 4
SQ SEQUENCE 4 AA; 420 MW; 6DD33DD6F0000000 CRC64;

Query Match 47.6%; Score 10; DB 1; Length 4;
Best Local Similarity 66.7%; Pred. No. 1.4e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          3 IAK 5
              :||
Db          1 MAK 3
```

RESULT 2
RM01_YEAST
ID RM01_YEAST STANDARD; PRT; 4 AA.
AC P36515;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Mitochondrial 60S ribosomal protein L1 (YmL1) (Fragment).
GN MRPL1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE.
RX MEDLINE=91285106; PubMed=2060626;
RA Grohmann L., Graack H.-R., Kruft V., Choli T., Goldschmidt-Reisin S.,
RA Kitakawa M.;
RT "Extended N-terminal sequencing of proteins of the large ribosomal
RT subunit from yeast mitochondria.";
RL FEBS Lett. 284:51-56(1991).

```
DR   PIR; S17255; S17255.
DR   SGD; L0002681; MRPL1.
KW   Ribosomal protein; Mitochondrion.
FT   NON_TER       4      4
SQ   SEQUENCE   4 AA;  402 MW;  7771B2D5D0000000 CRC64;

  Query Match             38.1%;  Score 8;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SV 2
              ||
Db          1 SV 2


RESULT 3
UF01_MOUSE
ID   UF01_MOUSE     STANDARD;      PRT;     5 AA.
AC   P38639;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Unknown protein from 2D-page of fibroblasts (P19) (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Fibroblast;
RX   MEDLINE=95009907; PubMed=7523108;
RA   Merrick B.A., Patterson R.M., Wichter L.L., He C., Selkirk J.K.;
RT   "Separation and sequencing of familiar and novel murine proteins
RT   using preparative two-dimensional gel electrophoresis.";
RL   Electrophoresis 15:735-745(1994).
CC   -!- MISCELLANEOUS: On the 2D-gel the determined pI of this unknown
CC       protein is: 6.6, its MW is: 19 kDa.
FT   NON_TER       5      5
SQ   SEQUENCE   5 AA;  717 MW;  7364087043100000 CRC64;

  Query Match             28.6%;  Score 6;  DB 1;  Length 5;
  Best Local Similarity   33.3%;  Pred. No. 1.4e+05;
  Matches    1;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 IAK 5
              | :
Db          2 IGR 4


RESULT 4
GRWM_HUMAN
ID   GRWM_HUMAN     STANDARD;      PRT;     3 AA.
AC   P01157;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   21-JUL-1986 (Rel. 01, Last annotation update)
DE   Growth-modulating peptide.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=77162369; PubMed=858356;
RA   Schlesinger D.H., Pickart L., Thaler M.M.;
RT   "Growth-modulating serum tripeptide is glycyl-histidyl-lysine.";
RL   Experientia 33:324-325(1977).
CC   -!- MISCELLANEOUS: This serum tripeptide has been found to stimulate
CC       growth of some cell types and to inhibit other types in vitro.
DR   GO; GO:0001558; P:regulation of cell growth; NAS.
SQ   SEQUENCE   3 AA;  340 MW;  6331E81000000000 CRC64;

  Query Match             23.8%;  Score 5;  DB 1;  Length 3;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+05;
  Matches    1;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 K 5
              |
Db          3 K 3


RESULT 5
LUXE_VIBFI
ID   LUXE_VIBFI     STANDARD;      PRT;     3 AA.
AC   P24272;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Long-chain-fatty-acid--luciferin-component ligase (EC 6.2.1.19) (Acyl-
DE   protein synthetase) (Fragment).
GN   LUXE.
OS   Vibrio fischeri.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=668;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91072226; PubMed=2254256;
RA   Swartzman E., Kapoor S., Graham A.F., Meighen E.A.;
RT   "A new Vibrio fischeri lux gene precedes a bidirectional termination
RT   site for the lux operon.";
RL   J. Bacteriol. 172:6797-6802(1990).
CC   -!- FUNCTION: ACYL-PROTEIN SYNTHETASE ACTIVATES TETRADECANOIC ACID.
CC       IT IS A COMPONENT OF THE FATTY ACID REDUCTASE COMPLEX RESPONSIBLE
CC       FOR CONVERTING TETRADECANOIC ACID TO THE ALDEHYDE WHICH SERVES AS
CC       SUBSTRATE IN THE LUCIFERASE-CATALYZED REACTION.
CC   -!- CATALYTIC ACTIVITY: ATP + an acid + protein = AMP + diphosphate +
CC       an acyl-protein thiolester.
CC   -!- PATHWAY: Bioluminescent fatty acid reduction system; second step.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M62812; -; NOT_ANNOTATED_CDS.
KW   Luminescence; Ligase.
FT   NON_TER       1      1
SQ   SEQUENCE   3 AA;  374 MW;  6AA3303000000000 CRC64;

  Query Match             23.8%;  Score 5;  DB 1;  Length 3;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+05;
  Matches    1;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 K 5
              |
Db          2 K 2


RESULT 6
FFKA_ANTEL
ID   FFKA_ANTEL     STANDARD;      PRT;     4 AA.
AC   P58705;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Antho-KAamide.
OS   Anthopleura elegantissima (Sea anemone).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC   Nynantheae; Actiniidae; Anthopleura.
OX   NCBI_TaxID=6110;
RN   [1]
RP   SEQUENCE.
```

```
RX   MEDLINE=92028852; PubMed=1681803;
RA   Nothacker H.-P., Rinehart K.L. Jr., Grimmelikhuijzen C.J.P.;
RT   "Isolation of L-3-phenyllactyl-Phe-Lys-Ala-NH2 (Antho-KAamide), a
RT   novel neuropeptide from sea anemones.";
RL   Biochem. Biophys. Res. Commun. 179:1205-1211(1991).
RN   [2]
RP   FUNCTION.
RX   MEDLINE=93391436; PubMed=8397415;
RA   McFarlane I.D., Hudman D., Nothacker H.-P., Grimmelikhuijzen C.J.P.;
RT   "The expansion behaviour of sea anemones may be coordinated by two
RT   inhibitory neuropeptides, Antho-KAamide and Antho-RIamide.";
RL   Proc. R. Soc. Lond., B, Biol. Sci. 253:183-188(1993).
CC   -!- FUNCTION: Inhibits spontaneous contractions in several muscle
CC       groups. May be involved in the expansion phase of feeding
CC       behaviour in sea anemones.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Neuron specific.
DR   PIR; JQ1273; JQ1273.
KW   Neuropeptide; Amidation.
FT   MOD_RES       1       1       L-3-PHENYLLACTYL.
FT   MOD_RES       4       4       AMIDATION.
SQ   SEQUENCE   4 AA;  512 MW;  6DD339C9A0000000 CRC64;

  Query Match             23.8%;  Score 5;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+05;
  Matches    1;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 K 5
               |
Db          3 K 3


RESULT 7
TUFT_HUMAN
ID   TUFT_HUMAN     STANDARD;      PRT;     4 AA.
AC   P01858;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phagocytosis-stimulating peptide (Tuftsin).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=72187087; PubMed=4112769;
RA   Nishioka K., Constantopoulos A., Satoh P.S., Najjar V.A.;
RT   "The characteristics, isolation and synthesis of the phagocytosis
RT   stimulating peptide tuftsin.";
RL   Biochem. Biophys. Res. Commun. 47:172-179(1972).
RN   [2]
RP   IMMUNOGLOBULIN CLASS.
RX   MEDLINE=68091045; PubMed=4169272;
RA   Fidalgo B.V., Najjar V.A.;
RT   "The physiological role of the lymphoid system. VI. The stimulatory
RT   effect of leucophilic gamma globulin (leucokinin) on the phagocytic
RT   activity of human polymorphonuclear leucocyte.";
RL   Biochemistry 6:3386-3392(1967).
CC   -!- MISCELLANEOUS: An IgG (called leucokinin) binds reversibly to the
CC       cell membrane of neutrophils in the blood. Leucokininase on the
CC       membrane releases the active peptide tuftsin from the gamma chain.
CC       Tuftsin is essential for maximum stimulation of the phagocytic
CC       activity of neutrophils.
DR   PIR; A02147; A02147.
DR   MIM; 191150; -.
DR   GO; GO:0003823; F:antigen binding; NAS.
DR   GO; GO:0006909; P:phagocytosis; NAS.
SQ   SEQUENCE   4 AA;  501 MW;  74176321C0000000 CRC64;

  Query Match             23.8%;  Score 5;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+05;
  Matches    1;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 K 5
               |
Db          2 K 2


RESULT 8
BIOB_CITFR
ID   BIOB_CITFR     STANDARD;      PRT;     5 AA.
AC   P12997;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Biotin synthase (EC 2.8.1.6) (Biotin synthetase) (Fragment).
GN   BIOB.
OS   Citrobacter freundii.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Citrobacter.
OX   NCBI_TaxID=546;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89006280; PubMed=2971595;
RA   Shiuan D., Campbell A.;
RT   "Transcriptional regulation and gene arrangement of Escherichia coli,
RT   Citrobacter freundii and Salmonella typhimurium biotin operons.";
RL   Gene 67:203-211(1988).
CC   -!- CATALYTIC ACTIVITY: Dethiobiotin + sulfur = biotin.
CC   -!- PATHWAY: Biotin biosynthesis; last step.
CC   -!- SIMILARITY: Belongs to the biotin and lipoic acid synthetases
CC       family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M21922; -; NOT_ANNOTATED_CDS.
DR   PIR; I40698; I40698.
KW   Biotin biosynthesis; Iron-sulfur; Transferase.
FT   NON_TER       5       5
SQ   SEQUENCE   5 AA;  532 MW;  75A5B1EDD6F00000 CRC64;

  Query Match             23.8%;  Score 5;  DB 1;  Length 5;
  Best Local Similarity   50.0%;  Pred. No. 1.4e+05;
  Matches    1;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 IA 4
               :|
Db          1 MA 2


RESULT 9
BPP7_BOTIN
ID   BPP7_BOTIN     STANDARD;      PRT;     5 AA.
AC   P30425;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Bradykinin-potentiating peptide S5,2 (5A) (Angiotensin-converting
DE   enzyme inhibitor).
OS   Bothrops insularis (Island jararaca) (Queimada jararaca).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Bothrops.
OX   NCBI_TaxID=8723;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Venom;
RX   MEDLINE=90351557; PubMed=2386615;
RA   Cintra A.C.O., Vieira C.A., Giglio J.R.;
```

```
RT   "Primary structure and biological activity of bradykinin potentiating
RT   peptides from Bothrops insularis snake venom.";
RL   J. Protein Chem. 9:221-227(1990).
CC   -!- FUNCTION: This peptide both inhibits the activity of the
CC       angiotensin-converting enzyme and enhances the action of
CC       bradykinin by inhibiting the kinases that inactivate it.
CC       It acts as an indirect hypotensive agent.
DR   PIR; G37196; G37196.
KW   Hypotensive agent; Pyrrolidone carboxylic acid.
FT   MOD_RES       1       1       PYRROLIDONE CARBOXYLIC ACID.
SQ   SEQUENCE   5 AA;  629 MW;  776DC37326B00000 CRC64;

  Query Match             23.8%;  Score 5;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+05;
  Matches    1;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 K 5
              |
Db          2 K 2


RESULT 10
EI04_LITRU
ID   EI04_LITRU     STANDARD;      PRT;      5 AA.
AC   P82100;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Electrin 4.
OS   Litoria rubella (Desert tree frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Hyloidea; Hylidae;
OC   Pelodryadinae; Litoria.
OX   NCBI_TaxID=104895;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Skin secretion;
RA   Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;
RT   "Peptides from the skin glands of the Australian buzzing tree frog
RT   Litori electrica. Comparison with the skin peptides from Litoria
RT   rubella.";
RL   Aust. J. Chem. 52:639-645(1999).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Skin.
KW   Amphibian defense peptide; Amidation.
FT   MOD_RES       5       5       AMIDATION.
SQ   SEQUENCE   5 AA;  616 MW;  61F2D1A059A00000 CRC64;

  Query Match             23.8%;  Score 5;  DB 1;  Length 5;
  Best Local Similarity   50.0%;  Pred. No. 1.4e+05;
  Matches    1;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SV 2
              :|
Db          3 TV 4


RESULT 11
TRM3_ECOLI
ID   TRM3_ECOLI     STANDARD;      PRT;      5 AA.
AC   P13973;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   TraM protein (Fragment).
GN   TRAM.
OS   Escherichia coli.
OG   Plasmid IncFII R100.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88227859; PubMed=2836369;
RA   Inamoto S., Yoshioka Y., Ohtsubo E.;
RT   "Identification and characterization of the products from the traJ
RT   and traY genes of plasmid R100.";
RL   J. Bacteriol. 170:2749-2757(1988).
CC   -!- FUNCTION: TRANSFER GENE PROTEIN. IS INVOLVED IN THE CONJUGATION
CC       PROCESS OF BACTERIAL CELLS FOR THE EXCHANGE OF PLASMID DNA.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the traM family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M20941; -; NOT_ANNOTATED_CDS.
DR   PIR; A32014; A32014.
KW   Conjugation; Plasmid; DNA-binding.
FT   NON_TER       1       1
SQ   SEQUENCE   5 AA;  634 MW;  6B1B1AA443500000 CRC64;

  Query Match             23.8%;  Score 5;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+05;
  Matches    1;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 K 5
              |
Db          1 K 1


RESULT 12
ACH1_ACHFU
ID   ACH1_ACHFU     STANDARD;      PRT;      4 AA.
AC   P35904;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Achatin-I.
OS   Achatina fulica (Giant African snail).
OC   Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC   Sigmurethra; Achatinoidea; Achatinidae; Achatina.
OX   NCBI_TaxID=6530;
RN   [1]
RP   SEQUENCE, CHARACTERIZATION, AND SYNTHESIS.
RC   STRAIN=Ferussac; TISSUE=Ganglion;
RX   MEDLINE=89273551; PubMed=2597281;
RA   Kamatani Y., Minakata H., Kenny P.T.M., Iwashita T., Watanabe K.,
RA   Funase K., Sun X.P., Yongsiri A., Kim K.H., Novales-Li P.,
RA   Novales E.T., Kanapi C.G., Takeuchi H., Nomoto K.;
RT   "Achatin-I, an endogenous neuroexcitatory tetrapeptide from Achatina
RT   fulica Ferussac containing a D-amino acid residue.";
RL   Biochem. Biophys. Res. Commun. 160:1015-1020(1989).
RN   [2]
RP   CHARACTERIZATION.
RC   STRAIN=Ferussac; TISSUE=Heart atrium;
RX   MEDLINE=91264856; PubMed=1675568;
RA   Fujimoto K., Kubota I., Yasuda-Kamatani Y., Minakata H., Nomoto K.,
RA   Yoshida M., Harada A., Muneoka Y., Kobayashi M.;
RT   "Purification of achatin-I from the atria of the African giant snail,
RT   Achatina fulica, and its possible function.";
RL   Biochem. Biophys. Res. Commun. 177:847-853(1991).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY.
RX   MEDLINE=93014529; PubMed=1399265;
RA   Ishida T., In Y., Doi M., Inoue M., Yasuda-Kamatani Y., Minakata H.,
RA   Iwashita T., Nomoto K.;
RT   "Crystal structure and molecular conformation of achatin-I
RT   (H-Gly-D-Phe-Ala-Asp-OH), an endogenous neuropeptide containing a
RT   D-amino acid residue.";
RL   Int. J. Pept. Protein Res. 39:258-264(1992).
```

```
CC   -!- FUNCTION: Neuroexcitatory peptide; increases the impulse frequency
CC       and produces a spike broadening of the identified heart excitatory
CC       neuron (PON); also enhances the amplitude and frequency of the
CC       heart beat. Has also an effect on several other muscles.
DR   PIR; A32480; A32480.
KW   Hormone; D-amino acid.
FT   MOD_RES       2      2       D-PHENYLALANINE.
SQ   SEQUENCE   4 AA;  408 MW;  6AADD9C810000000 CRC64;

  Query Match             19.0%;  Score 4;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+05;
  Matches    1;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 A 4
              |
Db          3 A 3


RESULT 13
EOSI_HUMAN
ID   EOSI_HUMAN     STANDARD;      PRT;     4 AA.
AC   P02731;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   21-JUL-1986 (Rel. 01, Last annotation update)
DE   Eosinophilotactic peptides.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=76078412; PubMed=1060093;
RA   Goetzl E.J., Austen K.F.;
RT   "Purification and synthesis of eosinophilotactic tetrapeptides of
RT   human lung tissue: identification as eosinophil chemotactic factor of
RT   anaphylaxis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 72:4123-4127(1975).
CC   -!- MISCELLANEOUS: These peptides are released from mast cells in lung
CC       (and other tissues) during hypersensitivity reactions
CC       (anaphylaxis). Their activities, preferentially affecting
CC       eosinophils, include chemotaxis, chemotactic deactivation, release
CC       of enzymes, and stimulation of the hexose monophosphate shunt.
DR   GO; GO:0006935; P:chemotaxis; IDA.
DR   GO; GO:0006955; P:immune response; IDA.
FT   VARIANT       1      1       V -> A (IN OTHER PEPTIDE).
FT                                 /FTId=VAR_005201.
SQ   SEQUENCE   4 AA;  390 MW;  6B05B862A0000000 CRC64;

  Query Match             19.0%;  Score 4;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+05;
  Matches    1;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 V 2
              |
Db          1 V 1


RESULT 14
FYRI_ANTEL
ID   FYRI_ANTEL     STANDARD;      PRT;     4 AA.
AC   P58706;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Antho-RIamide I [Contains: Antho-RIamide II].
OS   Anthopleura elegantissima (Sea anemone).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC   Nynantheae; Actiniidae; Anthopleura.
OX   NCBI_TaxID=6110;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=92270459; PubMed=1821096;
RA   Nothacker H.-P., Rinehart K.L. Jr., McFarlane I.D.,
RA   Grimmelikhuijzen C.J.P.;
RT   "Isolation of two novel neuropeptides from sea anemones: the unusual,
RT   biologically active L-3-phenyllactyl-Tyr-Arg-Ile-NH2 and its
RT   des-phenyllactyl fragment Tyr-Arg-Ile-NH2.";
RL   Peptides 12:1165-1173(1991).
RN   [2]
RP   FUNCTION.
RX   MEDLINE=93391436; PubMed=8397415;
RA   McFarlane I.D., Hudman D., Nothacker H.-P., Grimmelikhuijzen C.J.P.;
RT   "The expansion behaviour of sea anemones may be coordinated by two
RT   inhibitory neuropeptides, Antho-KAamide and Antho-RIamide.";
RL   Proc. R. Soc. Lond., B, Biol. Sci. 253:183-188(1993).
CC   -!- FUNCTION: Inhibits spontaneous contractions in several muscle
CC       groups. May be involved in the expansion phase of feeding
CC       behaviour in sea anemones.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Neuron specific.
KW   Neuropeptide; Amidation.
FT   CHAIN         1      4       ANTHO-RIAMIDE I.
FT   CHAIN         2      4       ANTHO-RIAMIDE II.
FT   MOD_RES       1      1       L-3-PHENYLLACTYL.
FT   MOD_RES       4      4       AMIDATION.
SQ   SEQUENCE   4 AA;  598 MW;  60441B59A0000000 CRC64;

  Query Match             19.0%;  Score 4;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+05;
  Matches    1;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 I 3
              |
Db          4 I 4


RESULT 15
OCP3_OCTMI
ID   OCP3_OCTMI     STANDARD;      PRT;     4 AA.
AC   P58649;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cardioactive peptides Ocp-3/Ocp-4.
OS   Octopus minor (Octopus).
OC   Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC   Octopodiformes; Octopoda; Incirrata; Octopodidae; Octopus.
OX   NCBI_TaxID=89766;
RN   [1]
RP   SEQUENCE, SYNTHESIS, MASS SPECTROMETRY, AND CHARACTERIZATION.
RC   TISSUE=Brain;
RX   MEDLINE=20336815; PubMed=10876044;
RA   Iwakoshi E., Hisada M., Minakata H.;
RT   "Cardioactive peptides isolated from the brain of a Japanese octopus,
RT   Octopus minor.";
RL   Peptides 21:623-630(2000).
CC   -!- FUNCTION: Cardioactive; has both positive chronotropic and
CC       inotropic effects on the heart. Ocp-4 is a 1000 time less
CC       active than Ocp-3.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PTM: Ocp-4 has D-Ser instead of L-Ser.
CC   -!- MASS SPECTROMETRY: MW=395.2; METHOD=MALDI.
KW   Hormone; D-amino acid.
FT   MOD_RES       2      2       D-SERINE (IN OCP-4).
SQ   SEQUENCE   4 AA;  463 MW;  6AB365B810000000 CRC64;

  Query Match             19.0%;  Score 4;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+05;
  Matches    1;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 S 1
              |
Db          2 S 2
```

Search completed: August 12, 2004, 06:53:27
Job time : 14 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:32:33 ; Search time 75 Seconds
(without alignments)
21.035 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 6

Minimum DB seq length: 0
Maximum DB seq length: 5

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7	33.3	4	11	Q08433	Q08433 rattus sp.
2	5	23.8	4	5	P83568	P83568 sepia offic
3	5	23.8	5	2	P83073	P83073 bacillus ce
4	5	23.8	5	10	Q99007	Q99007 hordeum vul
5	2	9.5	5	13	P83308	P83308 gallus gall
6	0	0.0	2	5	P83570	P83570 sepia offic

ALIGNMENTS

RESULT 1
Q08433
ID Q08433 PRELIMINARY; PRT; 4 AA.
AC Q08433;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Bilirubin UDP-glucuronosyltransferase (Fragment).
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10118;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Gunn;
RX MEDLINE=91282758; PubMed=1840486;
RA Sato H., Aono S., Kashiwamata S., Koiwai O.;
RT "Genetic defect of bilirubin UDP-glucuronosyltransferase in the
RT hyperbilirubinemic Gunn rat.";
RL Biochem. Biophys. Res. Commun. 177:1161-1164(1991).
DR EMBL; S38636; AAB19259.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.
KW Transferase.
FT NON_TER 1 1
SQ SEQUENCE 4 AA; 473 MW; 633732C420000000 CRC64;

Query Match 33.3%; Score 7; DB 11; Length 4;
Best Local Similarity 33.3%; Pred. No. 1e+06;
Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVI 3
:|:
Db 1 NVL 3

RESULT 2
P83568
ID P83568 PRELIMINARY; PRT; 4 AA.
AC P83568;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Pheromone peptide ILME.
OS Sepia officinalis (Common cuttlefish).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC Decapodiformes; Sepioidea; Sepiidae; Sepia.
OX NCBI_TaxID=6610;
RN [1]
RP SEQUENCE, SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND MASS
RP SPECTROMETRY.
RC TISSUE=Egg;
RX PubMed=10944467;
RA Zatylny C., Gagnon J., Boucaud-Camou E., Henry J.;
RT "ILME: a waterborne pheromonal peptide released by the eggs of Sepia
RT officinalis.";
RL Biochem. Biophys. Res. Commun. 275:217-222(2000).
RN [2]
RP SEQUENCE.
RC TISSUE=Egg;
RX PubMed=12207899;
RA Zatylny C., Marvin L., Gagnon J., Henry J.;
RT "Fertilization in Sepia officinalis: the first mollusk sperm-
RT attracting peptide.";
RL Biochem. Biophys. Res. Commun. 296:1186-1193(2002).
CC -!- FUNCTION: HAS MYOTROPIC ACTIVITY TARGETING THE GENITAL TRACT.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- TISSUE SPECIFICITY: FOLLICLE, FULLY GROWN OOCYTE AND EGG(EC2).
CC -!- MASS SPECTROMETRY: MW=505.4; METHOD=MALDI.
DR GO; GO:0005186; F:pheromone activity; IEA.
KW Pheromone.
SQ SEQUENCE 4 AA; 505 MW; 6B16972030000000 CRC64;

Query Match 23.8%; Score 5; DB 5; Length 4;
Best Local Similarity 0.0%; Pred. No. 1e+06;
Matches 0; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VI 3
::

```
Db          1 IL 2


RESULT 3
P83073
ID   P83073      PRELIMINARY;      PRT;     5 AA.
AC   P83073;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   88 kDa protein (Fragment).
OS   Bacillus cereus.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1396;
RN   [1]
RP   SEQUENCE.
RC   STRAIN=NCIMB 11796;
RA   Browne N., Dowds B.C.A.;
RL   Submitted (JUL-2001) to Swiss-Prot.
FT   NON_TER       5       5
SQ   SEQUENCE   5 AA;  623 MW;  6B01AAA336F00000 CRC64;

  Query Match             23.8%;  Score 5;  DB 2;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1e+06;
  Matches    1;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 K 5
              |
Db          2 K 2


RESULT 4
Q99007
ID   Q99007      PRELIMINARY;      PRT;     5 AA.
AC   Q99007;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Alpha amylase (Fragment).
GN   AMY1 GENE.
OS   Hordeum vulgare (Barley).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Triticeae; Hordeum.
OX   NCBI_TaxID=4513;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91329704; PubMed=1831055;
RA   Jacobsen J.V., Close T.J.;
RT   "Control of transient expression of chimaeric genes by gibberellic
RT   acid and abscisic acid in protoplasts prepared from mature bareley
RT   aleurone layers.";
RL   Plant Mol. Biol. 16:713-721(1991).
DR   EMBL; X54643; CAA38455.1; -.
FT   NON_TER       5       5
SQ   SEQUENCE   5 AA;  600 MW;  61E3344DD6F00000 CRC64;

  Query Match             23.8%;  Score 5;  DB 10;  Length 5;
  Best Local Similarity   50.0%;  Pred. No. 1e+06;
  Matches    1;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 IA 4
              :|
Db          1 MA 2


RESULT 5
P83308
ID   P83308      PRELIMINARY;      PRT;     5 AA.
AC   P83308;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   FMRFamide-like neuropeptide (LPLRF-amide).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RC   TISSUE=Brain;
RX   PubMed=6137771;
RA   Dockray G.J., Reeve J.R. Jr., Shively J., Gayton R.J., Barnard C.S.;
RT   "A novel active pentapeptide from chicken brain identified by
RT   antibodies to FMRFamide.";
RL   Nature 305:328-330(1983).
CC   -!- FUNCTION: MAY FUNCTION AS A NEUROTRANSMITTER OR MODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   GO; GO:0007218; P:neuropeptide signaling pathway; TAS.
KW   Neuropeptide; Amidation.
FT   MOD_RES       5       5       AMIDATION.
SQ   SEQUENCE   5 AA;  645 MW;  69D4073767400000 CRC64;

  Query Match             9.5%;  Score 2;  DB 13;  Length 5;
  Best Local Similarity   0.0%;  Pred. No. 1e+06;
  Matches    0;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 I 3
              :
Db          1 L 1


RESULT 6
P83570
ID   P83570      PRELIMINARY;      PRT;     2 AA.
AC   P83570;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Neuropeptide GWa.
OS   Sepia officinalis (Common cuttlefish).
OC   Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC   Decapodiformes; Sepioidea; Sepiidae; Sepia.
OX   NCBI_TaxID=6610;
RN   [1]
RP   SEQUENCE, FUNCTION, MASS SPECTROMETRY, AND AMIDATION.
RC   TISSUE=Optic lobe;
RX   PubMed=9437704;
RA   Henry J., Favrel P., Boucaud-Camou E.;
RT   "Isolation and identification of a novel Ala-Pro-Gly-Trp-amide-related
RT   peptide inhibiting the motility of the mature oviduct in the
RT   cuttlefish, Sepia officinalis.";
RL   Peptides 18:1469-1474(1997).
CC   -!- FUNCTION: REGULATORY NEUROPEPTIDE WITH MYOTROPIC ACTIVITY
CC       TARGETING THE DISTAL OVIDUCT. INHIBITS THE MOTILITY OF THE OVIDUCT
CC       BY DECREASING TONUS, FREQUENCY AND AMPLITUDE OF CONTRACTIONS.
CC   -!- MASS SPECTROMETRY: MW=259.9; METHOD=MALDI.
DR   GO; GO:0007218; P:neuropeptide signaling pathway; IEA.
KW   Neuropeptide; Amidation.
FT   MOD_RES       2       2       AMIDATION.
SQ   SEQUENCE   2 AA;  261 MW;  7378100000000000 CRC64;

  Query Match             0.0%;  Score 0;  DB 5;  Length 2;
  Best Local Similarity   0.0%;  Pred. No. 1e+06;
  Matches    0;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 S 1

Db          1 G 1


Search completed: August 12, 2004, 06:54:52
Job time : 75 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:20:13 ; Search time 90 Seconds
(without alignments)
15.697 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 34717

Minimum DB seq length: 0
Maximum DB seq length: 5

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	100.0	5	3	AAY97147	Aay97147 Pigment p
2	21	100.0	5	5	ABB99061	Abb99061 N-termina
3	16	76.2	5	2	AAR72928	Aar72928 Yeast PPI
4	16	76.2	5	2	AAR72927	Aar72927 Yeast PPI
5	16	76.2	5	2	AAR72884	Aar72884 Yeast PPI
6	16	76.2	5	6	AAO29981	Aao29981 Peptide #
7	16	76.2	5	7	ABR63446	Abr63446 Rat purin
8	15	71.4	5	1	AAP97806	Aap97806 Sequence
9	14	66.7	4	2	AAR61324	Aar61324 Fragment
10	14	66.7	5	2	AAY07986	Aay07986 Human sec
11	13	61.9	4	1	AAP91629	Aap91629 Motif use
12	13	61.9	4	1	AAP97808	Aap97808 Sequence
13	13	61.9	4	2	AAW55770	Aaw55770 Immunisat
14	13	61.9	4	5	ABB84333	Abb84333 Human MBP
15	13	61.9	4	7	ABR57300	Abr57300 Thermus o
16	13	61.9	5	2	AAR12661	Aar12661 Pentapept
17	13	61.9	5	2	AAR51525	Aar51525 Mimotope
18	13	61.9	5	2	AAR69893	Aar69893 Pentameri
19	13	61.9	5	2	AAR78989	Aar78989 p107 pept
20	13	61.9	5	2	AAR98639	Aar98639 Peptide 1
21	12	57.1	4	2	AAR15757	Aar15757 Farnesyl-
22	12	57.1	4	2	AAR49753	Aar49753 Farnesylt
23	12	57.1	4	2	AAR77816	Aar77816 Farnesyl
24	12	57.1	4	2	AAW04443	Aaw04443 Farnesyl
25	12	57.1	4	2	AAW65412	Aaw65412 Peptide u
26	12	57.1	4	2	AAY28344	Aay28344 Peptide f
27	12	57.1	4	3	AAY87947	Aay87947 Mammalian
28	12	57.1	4	4	AAG65468	Aag65468 Substrate
29	12	57.1	4	4	AAB57512	Aab57512 Mannose r
30	12	57.1	4	4	AAB57922	Aab57922 Mannose r
31	12	57.1	4	4	AAB55665	Aab55665 Monocyte
32	12	57.1	4	4	AAB80566	Aab80566 Peptide u
33	12	57.1	4	5	ABJ05144	Abj05144 T-cell su
34	12	57.1	4	5	ABP63437	Abp63437 Monocyte
35	12	57.1	4	5	AAE20561	Aae20561 Soybean d
36	12	57.1	4	6	ABU79151	Abu79151 Prenylati
37	12	57.1	4	7	ADC26827	Adc26827 Anti-angi
38	12	57.1	4	7	ADD11758	Add11758 T cell su
39	12	57.1	5	1	AAP61368	Aap61368 N-termina
40	12	57.1	5	2	AAR11930	Aar11930 Pentapept
41	12	57.1	5	2	AAR51510	Aar51510 Mimotope
42	12	57.1	5	2	AAR71699	Aar71699 pBSmutLac
43	12	57.1	5	2	AAR71698	Aar71698 pBSmutLac
44	12	57.1	5	2	AAR69878	Aar69878 Pentameri
45	12	57.1	5	2	AAR66898	Aar66898 Agonist p

ALIGNMENTS

RESULT 1
AAY97147
ID AAY97147 standard; peptide; 5 AA.
XX
AC AAY97147;
XX
DT 04-DEC-2000 (first entry)
XX
DE Pigment protein from coral tissue N-terminal peptide 1.
XX
KW N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW UV filter.
XX
OS Acropora aspera.
OS Acropora horrida.
OS Montipora caliculata.
OS Montipora monasteriata.
OS Porites murrayensis.
OS Porites lobata.
XX
PN WO200046233-A1.
XX
PD 10-AUG-2000.
XX
PF 02-FEB-2000; 2000WO-AU000056.
XX
PR 02-FEB-1999; 99AU-00008463.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Hoegh-Guldberg O, Dove S;
XX
DR WPI; 2000-532892/48.
XX
PT Novel pigment protein derived from corals capable of emitting
PT fluorescence upon irradiation by incident light useful as tissue marker,
PT fluorescent marker or general dyestuff.
XX
PS Claim 3; Page 42; 49pp; English.
XX
CC The N-terminal peptides shown in AAY97147-48 are from pigment protein
CC from coral tissue (PPCT). PPCT is capable of emitting fluorescence upon
CC irradiation by incident light whose maximal absorbance is in the range of
CC 320-600 nm and a maximal fluorescence emission is in the range of 300-700
CC nm. PPCT may be used as a tissue marker, fluorescent marker (e.g. to
CC follow gene expression in transformed tissues) or general dyestuff (all
CC claimed). PPCT may also be used in sunscreen formulations or UV filters

```
CC   (both claimed)
XX
SQ   Sequence 5 AA;

  Query Match             100.0%;  Score 21;  DB 3;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db          1 SVIAK 5


RESULT 2
ABB99061
ID   ABB99061 standard; peptide; 5 AA.
XX
AC   ABB99061;
XX
DT   22-JAN-2003  (first entry)
XX
DE   N-terminal amino acid sequence of a CFM #1.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 3; Page 278; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 5 AA;

  Query Match             100.0%;  Score 21;  DB 5;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db          1 SVIAK 5


RESULT 3
AAR72928
ID   AAR72928 standard; peptide; 5 AA.
XX
AC   AAR72928;
XX
DT   16-OCT-2003  (revised)
DT   25-MAR-2003  (revised)
DT   29-NOV-1995  (first entry)
XX
DE   Yeast PPIase tryptic fragment 8.
XX
KW   Escherichia coli; protein conformation; folding; acceleration;
KW   PPIase-alpha; peptidyl prolyl cis trans isomerase alpha; catalysis;
KW   isomerisation; prolyl peptide bond.
XX
OS   Saccharomyces cerevisiae; strain AH22.
XX
PN   EP647714-A1.
XX
PD   12-APR-1995.
XX
PF   19-JUL-1990;   94EP-00203612.
XX
PR   19-JUL-1989;   89JP-00184738.
PR   06-OCT-1989;   89JP-00260244.
PR   29-DEC-1989;   89JP-00344705.
PR   19-JUL-1990;   90EP-00307914.
XX
PA   (TOFU ) TONEN CORP.
XX
PI   Hayano T,  Katou S,  Maki N,  Takahashi N,  Suzuki M;
XX
DR   WPI; 1995-140756/19.
XX
PT   New E.coli peptidyl prolyl cis trans isomerase beta - used to accelerate
PT   the folding of proteins, partic. for activation of inactive recombinant
PT   proteins.
XX
PS   Example 2; Page 23; 85pp; English.
XX
CC   AAR72921-29 are tryptic fragments of a yeast PPIase (peptidyl prolyl cis
CC   trans isomerase). The yeast PPIase has a single mol. wt. of about 17 kDa
CC   and a single isoelectric point of about 6.2. The enzyme catalyses the
CC   isomerisation of prolyl peptide bonds in proteins and accelerates the
CC   folding of the protein. The inventors are claiming a PPIase-beta.
CC   (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-MAR-2003 to
CC   correct PF field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated
CC   on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 5 AA;

  Query Match             76.2%;  Score 16;  DB 2;  Length 5;
  Best Local Similarity   75.0%;  Pred. No. 1.4e+06;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VIAK 5
```

```
              |:||
Db          2 VVAK 5


RESULT 4
AAR72927
ID   AAR72927 standard; peptide; 5 AA.
XX
AC   AAR72927;
XX
DT   16-OCT-2003  (revised)
DT   25-MAR-2003  (revised)
DT   29-NOV-1995  (first entry)
XX
DE   Yeast PPIase tryptic fragment 7.
XX
KW   Escherichia coli; protein conformation; folding; acceleration;
KW   PPIase-alpha; peptidyl prolyl cis trans isomerase alpha; catalysis;
KW   isomerisation; prolyl peptide bond.
XX
OS   Saccharomyces cerevisiae; strain AH22.
XX
PN   EP647714-A1.
XX
PD   12-APR-1995.
XX
PF   19-JUL-1990;   94EP-00203612.
XX
PR   19-JUL-1989;   89JP-00184738.
PR   06-OCT-1989;   89JP-00260244.
PR   29-DEC-1989;   89JP-00344705.
PR   19-JUL-1990;   90EP-00307914.
XX
PA   (TOFU ) TONEN CORP.
XX
PI   Hayano T,  Katou S,  Maki N,  Takahashi N,  Suzuki M;
XX
DR   WPI; 1995-140756/19.
XX
PT   New E.coli peptidyl prolyl cis trans isomerase beta - used to accelerate
PT   the folding of proteins, partic. for activation of inactive recombinant
PT   proteins.
XX
PS   Example 2; Page 23; 85pp; English.
XX
CC   AAR72921-29 are tryptic fragments of a yeast PPIase (peptidyl prolyl cis
CC   trans isomerase). The yeast PPIase has a single mol. wt. of about 17 kDa
CC   and a single isoelectric point of about 6.2. The enzyme catalyses the
CC   isomerisation of prolyl peptide bonds in proteins and accelerates the
CC   folding of the protein. The inventors are claiming a PPIase-beta.
CC   (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-MAR-2003 to
CC   correct PF field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated
CC   on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 5 AA;

  Query Match             76.2%;  Score 16;  DB 2;  Length 5;
  Best Local Similarity   75.0%;  Pred. No. 1.4e+06;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VIAK 5
              |:||
Db          2 VVAK 5


RESULT 5
AAR72884
ID   AAR72884 standard; peptide; 5 AA.
XX
AC   AAR72884;
XX
DT   16-OCT-2003  (revised)
DT   25-MAR-2003  (revised)
DT   29-NOV-1995  (first entry)
XX
DE   Yeast PPIase tryptic fragment 8.
XX
KW   Escherichia coli; protein conformation; folding; acceleration;
KW   PPIase-alpha; peptidyl prolyl cis trans isomerase alpha; catalysis;
KW   isomerisation; prolyl peptide bond.
XX
OS   Saccharomyces cerevisiae; strain AH22.
XX
PN   EP647713-A1.
XX
PD   12-APR-1995.
XX
PF   19-JUL-1990;   94EP-00203610.
XX
PR   19-JUL-1989;   89JP-00184738.
PR   06-OCT-1989;   89JP-00260244.
PR   29-DEC-1989;   89JP-00344705.
PR   19-JUL-1990;   90EP-00307914.
XX
PA   (TOFU ) TONEN CORP.
XX
PI   Hayano T,  Katou S,  Maki N,  Takahashi N,  Suzuki M;
XX
DR   WPI; 1995-140755/19.
XX
PT   New E.coli peptidyl prolyl cis trans isomerase alpha - used to accelerate
PT   the folding of proteins, partic. for activation of inactive recombinant
PT   proteins.
XX
PS   Example 2; Page 23; 85pp; English.
XX
CC   AAR72877-85 are tryptic fragments of a yeast PPIase (peptidyl prolyl cis
CC   trans isomerase). The yeast PPIase has a single mol. wt. of about 17 kDa
CC   and a single isoelectric point of about 6.2. The enzyme catalyses the
CC   isomerisation of prolyl peptide bonds in proteins and accelerates the
CC   folding of the protein. The inventors are claiming the PPIase-alpha.
CC   (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-MAR-2003 to
CC   correct PF field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated
CC   on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 5 AA;

  Query Match             76.2%;  Score 16;  DB 2;  Length 5;
  Best Local Similarity   75.0%;  Pred. No. 1.4e+06;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VIAK 5
              |:||
Db          2 VVAK 5


RESULT 6
AAO29981
ID   AAO29981 standard; peptide; 5 AA.
XX
AC   AAO29981;
XX
DT   03-SEP-2003  (first entry)
XX
DE   Peptide #4 used to enhance cellular adherence and growth.
XX
KW   Cellular adhesion; growth; expression; secretion.
XX
OS   Unidentified.
XX
PN   WO2003044045-A2.
XX
PD   30-MAY-2003.
XX
PF   19-NOV-2002; 2002WO-US037207.
XX
```

```
PR   19-NOV-2001; 2001US-00992124.
XX
PA   (BECT ) BECTON DICKINSON & CO.
XX
PI   Campbell RL,  Heidaran M,  Spargo CA,  Wilkins JH,  Haaland PD;
XX
DR   WPI; 2003-482396/45.
XX
PT   New peptide, useful for enhancing cell growth and/or secretion in a cell
PT   culture system.
XX
PS   Claim 3; Page 44; 78pp; English.
XX
CC   The invention relates to peptides having cell adhesion, growth,
CC   expression or secretion-enhancing activities. The peptides of the
CC   invention are useful for enhancing cell growth and/or secretion in a cell
CC   culture system. The present sequence is a peptide used to enhance
CC   cellular adherence and growth
XX
SQ   Sequence 5 AA;

  Query Match             76.2%;  Score 16;  DB 6;  Length 5;
  Best Local Similarity   75.0%;  Pred. No. 1.4e+06;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VIAK 5
              |:||
Db          2 VVAK 5


RESULT 7
ABR63446
ID   ABR63446 standard; peptide; 5 AA.
XX
AC   ABR63446;
XX
DT   08-SEP-2003  (first entry)
XX
DE   Rat purinergic receptor P2X3 peptide #2.
XX
KW   Polypeptide quantitation; amount of protein; rat; human;
KW   purinergic receptor; P2X3.
XX
OS   Rattus sp.
XX
PN   WO2003046148-A2.
XX
PD   05-JUN-2003.
XX
PF   27-NOV-2002; 2002WO-US038334.
XX
PR   29-NOV-2001; 2001US-0334325P.
XX
PA   (THER-) THERMO FINNIGAN LLC.
XX
PI   Barnidge DR,  Lindall AW;
XX
DR   WPI; 2003-505195/47.
XX
PT   Determining the actual amount of a polypeptide in a sample by measuring
PT   the amount of a cleavage product released from the polypeptide and using
PT   an exogenous polypeptide corresponding to the cleavage product as a
PT   standard.
XX
PS   Disclosure; Page 7; 26pp; English.
XX
CC   The present invention relates to a method of determining the actual
CC   amount of 1 or more selected polypeptides in a sample, which comprises
CC   releasing specific cleavage products from each of the polypeptides with
CC   cleavage agents and determining the actual amount of the product by
CC   comparison to a defined amount of a corresponding exogenous polypeptide
CC   (the actual amount of each of the cleavage products is directly related
CC   to the actual amount of the selected polypeptide from which it was
CC   released). The method is useful in quantifying the actual amount of 1 or
CC   more selected polypeptides in a sample, which may be important in
CC   research settings (e.g. in drug discovery and development) and in
CC   clinical settings (e.g. for medical diagnosis and for monitoring
CC   treatment efficacy). The present sequence is a peptide used to
CC   demonstrate the method of the invention
XX
SQ   Sequence 5 AA;

  Query Match             76.2%;  Score 16;  DB 7;  Length 5;
  Best Local Similarity   60.0%;  Pred. No. 1.4e+06;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              ||: |
Db          1 SVVVK 5


RESULT 8
AAP97806
ID   AAP97806 standard; protein; 5 AA.
XX
AC   AAP97806;
XX
DT   29-JUL-1992  (first entry)
XX
DE   Sequence of fragment 19, the tryptic fragment of recombinant penicillin
DE   acyltransferase (PAT) polypeptide 2.
XX
KW   Penicillin biosynthesis; enzyme; antibiotic.
XX
OS   Penicillium chrysogenum.
XX
PN   EP336446-A.
XX
PD   11-OCT-1989.
XX
PF   07-APR-1989;   89EP-00106214.
XX
PR   08-APR-1988;   88AT-00000922.
PR   13-JUL-1988;   88AT-00001806.
PR   08-SEP-1988;   88AT-00002201.
XX
PA   (BIOC ) BIOCHEMIE GMBH.
XX
PI   Knauseder F,  Leitner E,  Palma N,  Weber G;
XX
DR   WPI; 1989-294357/41.
XX
PT   Recombinant penicillin acyl-transferase - and DNA coding for it.
XX
PS   Claim 9; Page 48; 52pp; English.
XX
CC   The inventors claim recombinant penicillin acyltransferase (PAT) and DNA
CC   coding for PAT. PAT catalyses the last step in the biosynthesis of
CC   penicillin G and penicillin V. More specifically, the coding strand of
CC   the DNA has the nucleotide sequence shown below. This includes three
CC   introns and codes for a PAT protein with mol. wt. ca. 40kD. Plasmid
CC   vectors pBC2001 and pBC2002 are specifically claimed
XX
SQ   Sequence 5 AA;

  Query Match             71.4%;  Score 15;  DB 1;  Length 5;
  Best Local Similarity   60.0%;  Pred. No. 1.4e+06;
  Matches    3;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              :|||:
Db          1 AVIAR 5


RESULT 9
AAR61324
```

```
ID   AAR61324 standard; peptide; 4 AA.
XX
AC   AAR61324;
XX
DT   16-OCT-2003  (revised)
DT   25-MAR-2003  (revised)
DT   27-APR-1995  (first entry)
XX
DE   Fragment of deacetylase enzyme.
XX
KW   Deacetylase; enzyme; L-N-acetylphosphinothricin; L-AcPPT;
KW   L-phosphonothricin; PPT; glutamine synthase; plant; male sterility;
KW   anther.
XX
OS   Variovorax paradoxus; (mixed culture).
OS   Brevundimonas diminuta; (mixed culture).
OS   Nocardia globerula; (mixed culture).
OS   Cellulosimicrobium cellulans; (mixed culture).
OS   Agrobacterium tumefaciens; (mixed culture).
XX
PN   DE4308061-A1.
XX
PD   15-SEP-1994.
XX
PF   13-MAR-1993;   93DE-04308061.
XX
PR   13-MAR-1993;   93DE-04308061.
XX
PA   (FARH ) HOECHST AG.
XX
PI   Schulz A,  Bartsch K;
XX
DR   WPI; 1994-286683/36.
XX
PT   New de-acetylase specific for L-N-acetyl-phosphinothricin - isolated from
PT   soil microbes, useful for stereoselective prodn. of L-phosphinothricin
PT   and for making male-sterile plants.
XX
PS   Claim 5; Page 5; 5pp; German.
XX
CC   The deacetylase has a molecular weight of 20000-100000, an optimum pH of
CC   6.5-9.5 and substrate specificity for L-N-acetylphosphinothricin (L-
CC   AcPPT). It may be used for the deacetylation of AcPPT for the
CC   steroselective production of L-phosphinothricin (PPT) and for inducing
CC   reversible male serility in plants (PPT inhibits the enzyme glutamine
CC   synthase in anthers). The deacetylase comprises at least one of four
CC   sequences (See AAR61321-24). (Updated on 25-MAR-2003 to correct PN
CC   field.) (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 4 AA;

  Query Match             66.7%;  Score 14;  DB 2;  Length 4;
  Best Local Similarity   75.0%;  Pred. No. 1.4e+06;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VIAK 5
              |:||
Db          1 VMAK 4


RESULT 10
AAY07986
ID   AAY07986 standard; protein; 5 AA.
XX
AC   AAY07986;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Human secreted protein fragment #3 encoded from gene 41.
XX
KW   Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW   gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW   developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW   immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW   renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW   cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW   pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW   arthritis; malignancy; digestive; endocrine; infection.
XX
OS   Homo sapiens.
XX
PN   WO9918208-A1.
XX
PD   15-APR-1999.
XX
PF   01-OCT-1998;   98WO-US020775.
XX
PR   02-OCT-1997;   97US-0060833P.
PR   02-OCT-1997;   97US-0060836P.
PR   02-OCT-1997;   97US-0060837P.
PR   02-OCT-1997;   97US-0060838P.
PR   02-OCT-1997;   97US-0060839P.
PR   02-OCT-1997;   97US-0060843P.
PR   02-OCT-1997;   97US-0060862P.
PR   02-OCT-1997;   97US-0060866P.
PR   02-OCT-1997;   97US-0060874P.
PR   02-OCT-1997;   97US-0060880P.
PR   02-OCT-1997;   97US-0060884P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Duan DR,  Florence KA,  Rosen CA,  Ruben SM,  Greene JM,  Young P;
PI   Ferrie AM,  Yu G,  Janat F,  Ni J,  Carter KC,  Endress GA,  Feng P;
PI   Lafleur DW,  Shi Y;
XX
DR   WPI; 1999-264022/22.
XX
PT   New isolated human genes and the secreted polypeptides they encode.
XX
PS   Disclosure; Page 344; 368pp; English.
XX
CC   This invention describes novel isolated human genes and the secreted
CC   proteins they encode. The products of the invention are useful for
CC   preventing, treating or ameliorating medical conditions, e.g. by protein
CC   or gene therapy. Also pathological conditions can be diagnosed by
CC   determining the amount of the new polypeptides in a sample or by
CC   determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 101 polynucleotides, based on
CC   which tissues they are most highly expressed in, and include developing
CC   products for the diagnosis or treatment of cancer, tumours,
CC   neurodegenerative disorders, developmental abnormalities and fetal
CC   deficiencies, blood disorders, leukemias, diseases of the immune system,
CC   autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC   allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC   disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC   transplant rejection, disorders involving osteoclasts such as
CC   osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC   infections and AIDS. The human secreted proteins of the invention are
CC   represented in AAY07852-Y07993 and the encoding nucleic acids are
CC   represented in AAX37451-X37552
XX
SQ   Sequence 5 AA;

  Query Match             66.7%;  Score 14;  DB 2;  Length 5;
  Best Local Similarity   40.0%;  Pred. No. 1.4e+06;
  Matches    2;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |:::|
Db          1 SMVSK 5


RESULT 11
AAP91629
ID   AAP91629 standard; protein; 4 AA.
XX
```

AC AAP91629;
XX
DT 25-MAR-2003 (revised)
DT 09-JUL-1990 (first entry)
XX
DE Motif useful in tolerization alone or in association with epitopes to the
DE acetyl choline receptor.
XX
KW Autoantigen; MBP; myelin basic protein; transplantation antigen;
KW myasthenia gravis; myasthenics; Transplantation antigen.
XX
OS Synthetic.
XX
PN EP304279-A.
XX
PD 22-FEB-1989.
XX
PF 17-AUG-1988; 88EP-00307608.
XX
PR 17-AUG-1987; 87US-00086694.
XX
PA (STRD) UNIV LELAND STANFORD JUNIOR.
XX
PI Steinman L, Zamvil S;
XX
DR WPI; 1989-055696/08.
XX
PT Oligopeptide and polypeptide compsns. - based on the amino acid sequence
PT of an immunogen and used for modulating the immune system.
XX
PS Disclosure; Page; 7pp; English.
XX
CC Sequences will normally be part of 9-15 amino acid sequence, excluded as
CC motifs for immunisation but useful in tolerisation. . (Updated on 25-MAR-
CC 2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PA field.)
CC (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ Sequence 4 AA;

Query Match 61.9%; Score 13; DB 1; Length 4;
Best Local Similarity 50.0%; Pred. No. 1.4e+06;
Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VIAK 5
::||
Db 1 LVAK 4

RESULT 12
AAP97808
ID AAP97808 standard; protein; 4 AA.
XX
AC AAP97808;
XX
DT 29-JUL-1992 (first entry)
XX
DE Sequence of fragment 21, the tryptic fragment of recombinant penicillin
DE acyltransferase (PAT) polypeptide 2.
XX
KW Penicillin biosynthesis; enzyme; antibiotic.
XX
OS Penicillium chrysogenum.
XX
PN EP336446-A.
XX
PD 11-OCT-1989.
XX
PF 07-APR-1989; 89EP-00106214.
XX
PR 08-APR-1988; 88AT-00000922.
PR 13-JUL-1988; 88AT-00001806.
PR 08-SEP-1988; 88AT-00002201.
XX
PA (BIOC) BIOCHEMIE GMBH.
XX
PI Knauseder F, Leitner E, Palma N, Weber G;
XX
DR WPI; 1989-294357/41.
XX
PT Recombinant penicillin acyl-transferase - and DNA coding for it.
XX
PS Claim 9; Page 48; 52pp; English.
XX
CC The inventors claim recombinant penicillin acyltransferase (PAT) and DNA
CC coding for PAT. PAT catalyses the last step in the biosynthesis of
CC penicillin G and penicillin V. More specifically, the coding strand of
CC the DNA has the nucleotide sequence shown below. This includes three
CC introns and codes for a PAT protein with mol. wt. ca. 40kD. Plasmid
CC vectors pBC2001 and pBC2002 are specifically claimed
XX
SQ Sequence 4 AA;

Query Match 61.9%; Score 13; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IAK 5
|||
Db 2 IAK 4

RESULT 13
AAW55770
ID AAW55770 standard; peptide; 4 AA.
XX
AC AAW55770;
XX
DT 25-MAR-2003 (revised)
DT 08-JUL-1998 (first entry)
XX
DE Immunisation motif associated with AChR 1.
XX
KW Myelin basic protein; immunity; immune response; neurological; T-cell;
KW human; immunogen; B-cell; transplantation antigen; immunomodulator.
XX
OS Unidentified.
XX
PN EP805162-A1.
XX
PD 05-NOV-1997.
XX
PF 17-AUG-1988; 97EP-00106788.
XX
PR 17-AUG-1987; 87US-00086694.
PR 17-AUG-1988; 88EP-00307608.
XX
PA (STRD) UNIV LELAND STANFORD JUNIOR.
XX
PI Steinman L, Zamvil S;
XX
DR WPI; 1998-034664/04.
XX
PT Polypeptide comprising human myelin basic protein fragment - useful as
PT immuno modulator.
XX
PS Disclosure; Page 8; 8pp; English.
XX
CC The present sequence represents an immunisation motif normally excluded,
CC but which may be used with advantage for tolerisation by itself or in
CC conjunction with other epitope sequences from the present invention. The
CC present invention describes a polypeptide comprising a human myelin basic
CC protein (hMBP) fragment including P89-101 of hMBP, excluding native hMBP.
CC The term P89-101 is not defined but may be intended to mean amino acids
CC 89-101 of hMBP. The polypeptide can be used for tolerising a mammalian
CC host immune system comprising B and T cells to an immunogen of interest,
CC wherein said immunogen is restricted by a transplantation antigen of said

```
CC   host. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 25-MAR-
CC   2003 to correct PR field.)
XX
SQ   Sequence 4 AA;

  Query Match             61.9%;  Score 13;  DB 2;  Length 4;
  Best Local Similarity   50.0%;  Pred. No. 1.4e+06;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VIAK 5
              ::||
Db          1 LVAK 4


RESULT 14
ABB84333
ID   ABB84333 standard; peptide; 4 AA.
XX
AC   ABB84333;
XX
DT   17-OCT-2002  (first entry)
XX
DE   Human MBP protein derived peptide SEQ ID 33.
XX
KW   MBP; myelin basic protein; human; tolerance; immune system;
KW   multiple sclerosis; autoimmune response; autoimmune disease;
KW   immunosuppressive; neuroprotective.
XX
OS   Homo sapiens.
XX
PN   US2002076412-A1.
XX
PD   20-JUN-2002.
XX
PF   07-JUN-1995;   95US-00484409.
XX
PR   17-AUG-1987;   87US-00086694.
PR   12-JUL-1989;   89US-00379500.
PR   01-MAY-1990;   90US-00517245.
PR   01-MAY-1991;   91WO-US002991.
PR   30-APR-1992;   92US-00877444.
PR   21-MAY-1993;   93US-00066325.
PR   22-SEP-1993;   93US-00125407.
XX
PA   (STEI/) STEINMAN L.
PA   (ZAMV/) ZAMVIL S.
XX
PI   Steinman L,  Zamvil S;
XX
DR   WPI; 2002-598709/64.
XX
PT   Modulating or tolerizing the immune system, useful for treating multiple
PT   sclerosis, by administering a peptide derived from human myelin binding
PT   protein.
XX
PS   Disclosure; Page 14; 21pp; English.
XX
CC   This invention describes a novel method for modulating or tolerizing the
CC   immune system, and for treating multiple sclerosis comprising
CC   administering a peptide derived from hMBP (human myelin basic protein).
CC   The peptide induces an autoimmune response (T cell) to a self-antigen (or
CC   part of it), and binds to an MHC (major histocompatibility complex)
CC   antigen of a host susceptible to autoimmune diseases, i.e. competes with
CC   binding to MBP and inhibit proliferation of MBP-reactive cells. The
CC   peptide has immunosuppressive and neuroprotective activity. This sequence
CC   represents a peptide derived from the human MBP protein which can be used
CC   for tolerization
XX
SQ   Sequence 4 AA;

  Query Match             61.9%;  Score 13;  DB 5;  Length 4;
  Best Local Similarity   50.0%;  Pred. No. 1.4e+06;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VIAK 5
              ::||
Db          1 LVAK 4


RESULT 15
ABR57300
ID   ABR57300 standard; peptide; 4 AA.
XX
AC   ABR57300;
XX
DT   09-SEP-2003  (first entry)
XX
DE   Thermus oshimai nucleic acid polymerase peptide 704-707 SEQ ID NO:26.
XX
KW   Thermus oshimai; nucleic acid polymerase; enzyme; DNA sequencing;
KW   amplification; reverse transcription; RNA amplification;
KW   primer extension.
XX
OS   Thermus oshimai.
XX
PN   WO2003048310-A2.
XX
PD   12-JUN-2003.
XX
PF   22-NOV-2002; 2002WO-US037764.
XX
PR   30-NOV-2001; 2001US-0334798P.
XX
PA   (APPL-) APPLERA CORP.
XX
PI   Bolchakova E,  Rozzelle J;
XX
DR   WPI; 2003-505286/47.
XX
PT   New nucleic acid, useful for DNA sequencing or amplification, reverse
PT   transcription, RNA amplification or primer extension reactions.
XX
PS   Disclosure; Page 32; 64pp; English.
XX
CC   The present invention describes an isolated nucleic acid (I) encoding a
CC   nucleic acid polymerase or a derivative nucleic acid polymerase with a
CC   mutation that decreases 5-3' exonuclease activity or that reduces
CC   discrimination against dideoxynucleotide triphosphates. Also described:
CC   (1) a vector comprising the nucleic acid (I); (2) a host cell comprising
CC   the nucleic acid (I); (3) a nucleic acid polymerase or its derivative;
CC   (4) a kit comprising a container containing the nucleic acid polymerase
CC   of (3); (5) making the nucleic acid polymerase of (3); (6) synthesising a
CC   DNA; (7) thermocyclic amplification of nucleic acid; and (8) primer
CC   extending a DNA. The nucleic acid (I) is useful for DNA sequencing or
CC   amplification, reverse transcription, RNA amplification or primer
CC   extension reactions. The present sequence represents a Thermus oshimai
CC   nucleic acid polymerase peptide, which is given in the exemplification of
CC   the present invention
XX
SQ   Sequence 4 AA;

  Query Match             61.9%;  Score 13;  DB 7;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 IAK 5
              |||
Db          2 IAK 4


Search completed: August 12, 2004, 06:53:04
Job time : 93 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:54:59 ; Search time 41 Seconds
(without alignments)
38.284 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 17582

Minimum DB seq length: 0
Maximum DB seq length: 5

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	16	76.2	5	10	US-09-992-124A-4	Sequence 4, Appli
2	14	66.7	5	10	US-09-862-145A-4	Sequence 4, Appli
3	14	66.7	5	10	US-09-862-145A-8	Sequence 8, Appli
4	14	66.7	5	10	US-09-862-145A-12	Sequence 12, Appl
5	14	66.7	5	10	US-09-862-145A-16	Sequence 16, Appl
6	14	66.7	5	10	US-09-862-145A-20	Sequence 20, Appl
7	14	66.7	5	10	US-09-862-145A-24	Sequence 24, Appl
8	14	66.7	5	14	US-10-195-730-329	Sequence 329, App
9	14	66.7	5	15	US-10-402-029-4	Sequence 4, Appli
10	14	66.7	5	15	US-10-402-029-8	Sequence 8, Appli
11	14	66.7	5	15	US-10-402-029-12	Sequence 12, Appl
12	14	66.7	5	15	US-10-402-029-16	Sequence 16, Appl
13	14	66.7	5	15	US-10-402-029-20	Sequence 20, Appl
14	14	66.7	5	15	US-10-402-029-24	Sequence 24, Appl
15	14	66.7	5	16	US-10-285-108A-4	Sequence 4, Appli
16	14	66.7	5	16	US-10-285-108A-8	Sequence 8, Appli
17	14	66.7	5	16	US-10-285-108A-12	Sequence 12, Appl
18	14	66.7	5	16	US-10-285-108A-16	Sequence 16, Appl
19	14	66.7	5	16	US-10-285-108A-20	Sequence 20, Appl
20	14	66.7	5	16	US-10-285-108A-24	Sequence 24, Appl
21	13	61.9	4	8	US-08-484-409-33	Sequence 33, Appl
22	13	61.9	4	14	US-10-303-109A-26	Sequence 26, Appl
23	13	61.9	5	14	US-10-197-927-17	Sequence 17, Appl
24	12	57.1	4	9	US-09-797-543-3	Sequence 3, Appli
25	12	57.1	4	9	US-09-359-325A-31	Sequence 31, Appl
26	12	57.1	4	9	US-09-945-249-38	Sequence 38, Appl
27	12	57.1	4	9	US-09-870-759-163	Sequence 163, App
28	12	57.1	4	10	US-09-751-708A-163	Sequence 163, App
29	12	57.1	4	12	US-10-371-406B-3	Sequence 3, Appli
30	12	57.1	4	13	US-10-016-717-4	Sequence 4, Appli
31	12	57.1	4	13	US-10-033-026-11	Sequence 11, Appl
32	12	57.1	4	13	US-10-078-458-8	Sequence 8, Appli
33	12	57.1	4	14	US-10-087-905-6	Sequence 6, Appli
34	12	57.1	4	14	US-10-287-639-2	Sequence 2, Appli
35	12	57.1	4	14	US-10-087-942-6	Sequence 6, Appli
36	12	57.1	4	14	US-10-083-894-18	Sequence 18, Appl
37	12	57.1	4	14	US-10-083-894-42	Sequence 42, Appl
38	12	57.1	5	8	US-08-859-699-20	Sequence 20, Appl
39	12	57.1	5	9	US-09-764-246-5	Sequence 5, Appli
40	12	57.1	5	9	US-09-875-519A-29	Sequence 29, Appl
41	12	57.1	5	9	US-09-359-325A-30	Sequence 30, Appl
42	12	57.1	5	9	US-09-748-114-28	Sequence 28, Appl
43	12	57.1	5	9	US-09-945-249-37	Sequence 37, Appl
44	12	57.1	5	10	US-09-992-124A-3	Sequence 3, Appli
45	12	57.1	5	10	US-09-562-912-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-992-124A-4
; Sequence 4, Application US/09992124A
; Publication No. US20030162289A1
; GENERAL INFORMATION:
; APPLICANT: Heidaran, Mohammad A.
; APPLICANT: Haaland, Perry D.
; APPLICANT: Wilkins, Jamie H.
; APPLICANT: Spargo, Catherine A.
; APPLICANT: Campbell, Robert L.
; TITLE OF INVENTION: Peptides Promoting Cell Adherence, Growth and Secretion
; FILE REFERENCE: 102-410
; CURRENT APPLICATION NUMBER: US/09/992,124A
; CURRENT FILING DATE: 2002-05-21
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 5
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide selected for biological activity
US-09-992-124A-4

Query Match 76.2%; Score 16; DB 10; Length 5;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VIAK 5
|:||
Db 2 VVAK 5

RESULT 2
US-09-862-145A-4
; Sequence 4, Application US/09862145A
; Publication No. US20030138388A1
; GENERAL INFORMATION:

```
;  APPLICANT: Seiberg, Miri
;  APPLICANT:  Shapiro, Stanley
;  TITLE OF INVENTION: Peptides and the Use Thereof in Darkening the Skin
;  FILE REFERENCE: J&J-1991
;  CURRENT APPLICATION NUMBER: US/09/862,145A
;  CURRENT FILING DATE:  2001-05-21
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (5)..(5)
;   OTHER INFORMATION: C-terminal Amidation
US-09-862-145A-4

  Query Match             66.7%;  Score 14;  DB 10;  Length 5;
  Best Local Similarity   60.0%;  Pred. No. 1.2e+06;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |:| |
Db          1 SLIGK 5


RESULT 3
US-09-862-145A-8
; Sequence 8, Application US/09862145A
; Publication No. US20030138388A1
; GENERAL INFORMATION:
;  APPLICANT: Seiberg, Miri
;  APPLICANT:  Shapiro, Stanley
;  TITLE OF INVENTION: Peptides and the Use Thereof in Darkening the Skin
;  FILE REFERENCE: J&J-1991
;  CURRENT APPLICATION NUMBER: US/09/862,145A
;  CURRENT FILING DATE:  2001-05-21
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
US-09-862-145A-8

  Query Match             66.7%;  Score 14;  DB 10;  Length 5;
  Best Local Similarity   60.0%;  Pred. No. 1.2e+06;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |:| |
Db          1 SLIGK 5


RESULT 4
US-09-862-145A-12
; Sequence 12, Application US/09862145A
; Publication No. US20030138388A1
; GENERAL INFORMATION:
;  APPLICANT: Seiberg, Miri
;  APPLICANT:  Shapiro, Stanley
;  TITLE OF INVENTION: Peptides and the Use Thereof in Darkening the Skin
;  FILE REFERENCE: J&J-1991
;  CURRENT APPLICATION NUMBER: US/09/862,145A
;  CURRENT FILING DATE:  2001-05-21
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: Palmitoyl N-terminus
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (5)..(5)
;   OTHER INFORMATION: Amidated C-terminus
US-09-862-145A-12

  Query Match             66.7%;  Score 14;  DB 10;  Length 5;
  Best Local Similarity   60.0%;  Pred. No. 1.2e+06;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |:| |
Db          1 SLIGK 5


RESULT 5
US-09-862-145A-16
; Sequence 16, Application US/09862145A
; Publication No. US20030138388A1
; GENERAL INFORMATION:
;  APPLICANT: Seiberg, Miri
;  APPLICANT:  Shapiro, Stanley
;  TITLE OF INVENTION: Peptides and the Use Thereof in Darkening the Skin
;  FILE REFERENCE: J&J-1991
;  CURRENT APPLICATION NUMBER: US/09/862,145A
;  CURRENT FILING DATE:  2001-05-21
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: Palmitoyl N-terminus
US-09-862-145A-16

  Query Match             66.7%;  Score 14;  DB 10;  Length 5;
  Best Local Similarity   60.0%;  Pred. No. 1.2e+06;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |:| |
Db          1 SLIGK 5


RESULT 6
US-09-862-145A-20
; Sequence 20, Application US/09862145A
; Publication No. US20030138388A1
; GENERAL INFORMATION:
;  APPLICANT: Seiberg, Miri
;  APPLICANT:  Shapiro, Stanley
;  TITLE OF INVENTION: Peptides and the Use Thereof in Darkening the Skin
;  FILE REFERENCE: J&J-1991
;  CURRENT APPLICATION NUMBER: US/09/862,145A
;  CURRENT FILING DATE:  2001-05-21
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
;   LENGTH: 5
```

```
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: Stearatoyl N-terminus
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (5)..(5)
;   OTHER INFORMATION: Stearatoyl C-terminus
US-09-862-145A-20

  Query Match             66.7%;  Score 14;  DB 10;  Length 5;
  Best Local Similarity   60.0%;  Pred. No. 1.2e+06;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |:| |
Db          1 SLIGK 5


RESULT 7
US-09-862-145A-24
; Sequence 24, Application US/09862145A
; Publication No. US20030138388A1
; GENERAL INFORMATION:
;  APPLICANT: Seiberg, Miri
;  APPLICANT:  Shapiro, Stanley
;  TITLE OF INVENTION: Peptides and the Use Thereof in Darkening the Skin
;  FILE REFERENCE: J&J-1991
;  CURRENT APPLICATION NUMBER: US/09/862,145A
;  CURRENT FILING DATE:  2001-05-21
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: Stearatoyl N-terminus
US-09-862-145A-24

  Query Match             66.7%;  Score 14;  DB 10;  Length 5;
  Best Local Similarity   60.0%;  Pred. No. 1.2e+06;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |:| |
Db          1 SLIGK 5


RESULT 8
US-10-195-730-329
; Sequence 329, Application US/10195730
; Publication No. US20030144492A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et. al
;  TITLE OF INVENTION: 101 Human Secreted Proteins
;  FILE REFERENCE: PZ017P1
;  CURRENT APPLICATION NUMBER: US/10/195,730
;  CURRENT FILING DATE:  2002-07-16
;  PRIOR APPLICATION NUMBER: US/09/281,976
;  PRIOR FILING DATE: 1999-03-31
;  PRIOR APPLICATION NUMBER: 60/060,837
;  PRIOR FILING DATE: 1997-10-02
;  PRIOR APPLICATION NUMBER: 60/060,862
;  PRIOR FILING DATE: 1997-10-02
;  NUMBER OF SEQ ID NOS: 390
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 329
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-195-730-329

  Query Match             66.7%;  Score 14;  DB 14;  Length 5;
  Best Local Similarity   40.0%;  Pred. No. 1.2e+06;
  Matches    2;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |:::|
Db          1 SMVSK 5


RESULT 9
US-10-402-029-4
; Sequence 4, Application US/10402029
; Publication No. US20040005288A1
; GENERAL INFORMATION:
;  APPLICANT: Lin, Connie
;  APPLICANT:  Seiberg, Miri
;  TITLE OF INVENTION: Compositions for Darkening the Skin
;  FILE REFERENCE: J&J-2099
;  CURRENT APPLICATION NUMBER: US/10/402,029
;  CURRENT FILING DATE:  2003-03-28
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (5)..(5)
;   OTHER INFORMATION: C-terminal Amidation
US-10-402-029-4

  Query Match             66.7%;  Score 14;  DB 15;  Length 5;
  Best Local Similarity   60.0%;  Pred. No. 1.2e+06;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |:| |
Db          1 SLIGK 5


RESULT 10
US-10-402-029-8
; Sequence 8, Application US/10402029
; Publication No. US20040005288A1
; GENERAL INFORMATION:
;  APPLICANT: Lin, Connie
;  APPLICANT:  Seiberg, Miri
;  TITLE OF INVENTION: Compositions for Darkening the Skin
;  FILE REFERENCE: J&J-2099
;  CURRENT APPLICATION NUMBER: US/10/402,029
;  CURRENT FILING DATE:  2003-03-28
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
US-10-402-029-8

  Query Match             66.7%;  Score 14;  DB 15;  Length 5;
```

```
  Best Local Similarity   60.0%;  Pred. No. 1.2e+06;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |:| |
Db          1 SLIGK 5


RESULT 11
US-10-402-029-12
; Sequence 12, Application US/10402029
; Publication No. US20040005288A1
; GENERAL INFORMATION:
;  APPLICANT: Lin, Connie
;  APPLICANT:  Seiberg, Miri
;  TITLE OF INVENTION: Compositions for Darkening the Skin
;  FILE REFERENCE: J&J-2099
;  CURRENT APPLICATION NUMBER: US/10/402,029
;  CURRENT FILING DATE:  2003-03-28
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: Palmitoyl N-terminus
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (5)..(5)
;   OTHER INFORMATION: Amidated C-terminus
US-10-402-029-12

  Query Match             66.7%;  Score 14;  DB 15;  Length 5;
  Best Local Similarity   60.0%;  Pred. No. 1.2e+06;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |:| |
Db          1 SLIGK 5


RESULT 12
US-10-402-029-16
; Sequence 16, Application US/10402029
; Publication No. US20040005288A1
; GENERAL INFORMATION:
;  APPLICANT: Lin, Connie
;  APPLICANT:  Seiberg, Miri
;  TITLE OF INVENTION: Compositions for Darkening the Skin
;  FILE REFERENCE: J&J-2099
;  CURRENT APPLICATION NUMBER: US/10/402,029
;  CURRENT FILING DATE:  2003-03-28
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: Palmitoyl N-terminus
US-10-402-029-16

  Query Match             66.7%;  Score 14;  DB 15;  Length 5;
  Best Local Similarity   60.0%;  Pred. No. 1.2e+06;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |:| |
Db          1 SLIGK 5


RESULT 13
US-10-402-029-20
; Sequence 20, Application US/10402029
; Publication No. US20040005288A1
; GENERAL INFORMATION:
;  APPLICANT: Lin, Connie
;  APPLICANT:  Seiberg, Miri
;  TITLE OF INVENTION: Compositions for Darkening the Skin
;  FILE REFERENCE: J&J-2099
;  CURRENT APPLICATION NUMBER: US/10/402,029
;  CURRENT FILING DATE:  2003-03-28
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: Stearatoyl N-terminus
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (5)..(5)
;   OTHER INFORMATION: Amidated C-terminus
US-10-402-029-20

  Query Match             66.7%;  Score 14;  DB 15;  Length 5;
  Best Local Similarity   60.0%;  Pred. No. 1.2e+06;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |:| |
Db          1 SLIGK 5


RESULT 14
US-10-402-029-24
; Sequence 24, Application US/10402029
; Publication No. US20040005288A1
; GENERAL INFORMATION:
;  APPLICANT: Lin, Connie
;  APPLICANT:  Seiberg, Miri
;  TITLE OF INVENTION: Compositions for Darkening the Skin
;  FILE REFERENCE: J&J-2099
;  CURRENT APPLICATION NUMBER: US/10/402,029
;  CURRENT FILING DATE:  2003-03-28
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: Stearatoyl N-terminus
US-10-402-029-24

  Query Match             66.7%;  Score 14;  DB 15;  Length 5;
  Best Local Similarity   60.0%;  Pred. No. 1.2e+06;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy          1 SVIAK 5
              |:| |
Db          1 SLIGK 5


RESULT 15
US-10-285-108A-4
; Sequence 4, Application US/10285108A
; Publication No. US20040091449A1
; GENERAL INFORMATION:
;  APPLICANT: Lin, Connie
;  APPLICANT:  Wu, Jane
;  TITLE OF INVENTION: Compositions for Darkening the Skin and/or Hir
;  FILE REFERENCE: J&J-2172
;  CURRENT APPLICATION NUMBER: US/10/285,108A
;  CURRENT FILING DATE:  2002-10-31
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (5)..(5)
;   OTHER INFORMATION: C-terminal Amidation
US-10-285-108A-4

  Query Match             66.7%;  Score 14;  DB 16;  Length 5;
  Best Local Similarity   60.0%;  Pred. No. 1.2e+06;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |:| |
Db          1 SLIGK 5


Search completed: August 12, 2004, 07:02:19
Job time : 41 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:55:24 ; Search time 50 Seconds
(without alignments)
96.066 Million cell updates/sec

Title: US-09-890-463-2
Perfect score: 83
Sequence: 1 SVIAKQMTYKVYMSGTV 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 470470

Minimum DB seq length: 0
Maximum DB seq length: 17

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	83	100.0	17	3	AAY97148	Aay97148 Pigment p
2	73	88.0	16	5	ABB99066	Abb99066 N-termina
3	72	86.7	16	5	ABB99073	Abb99073 N-termina
4	70	84.3	16	5	ABB99072	Abb99072 N-termina
5	69	83.1	16	5	ABB99068	Abb99068 N-termina
6	69	83.1	16	5	ABB99070	Abb99070 N-termina
7	68	81.9	16	5	ABB99067	Abb99067 N-termina
8	67	80.7	16	5	ABB99071	Abb99071 N-termina
9	66	79.5	16	5	ABB99069	Abb99069 N-termina
10	62	74.7	16	5	ABB99074	Abb99074 N-termina
11	36	43.4	13	5	ABP70008	Abp70008 Colour Fa
12	29	34.9	14	2	AAR77526	Aar77526 p45 metal
13	29	34.9	14	2	AAW05846	Aaw05846 Fusarium
14	29	34.9	15	5	AAM48968	Aam48968 Human zin
15	28	33.7	11	2	AAW39598	Aaw39598 Human mel
16	28	33.7	12	4	AAB45642	Aab45642 Vasoactiv
17	28	33.7	13	2	AAR69362	Aar69362 Stearoyl-
18	28	33.7	13	4	AAB45641	Aab45641 Vasoactiv
19	28	33.7	13	4	AAB45639	Aab45639 Vasoactiv
20	28	33.7	13	5	AAE19614	Aae19614 Human ste
21	28	33.7	14	2	AAR79549	Aar79549 Analgesic
22	28	33.7	14	4	AAB88179	Aab88179 CD66 pept
23	28	33.7	14	4	AAB45638	Aab45638 Vasoactiv
24	28	33.7	14	4	AAB45622	Aab45622 Vasoactiv
25	28	33.7	15	2	AAR79548	Aar79548 Analgesic
26	28	33.7	15	4	AAB99955	Aab99955 Human lat
27	28	33.7	15	4	AAB45619	Aab45619 Vasoactiv
28	28	33.7	15	4	AAB45621	Aab45621 Vasoactiv
29	28	33.7	15	5	ABG71317	Abg71317 Human Sai
30	28	33.7	16	2	AAR79547	Aar79547 Analgesic
31	28	33.7	16	3	AAY85708	Aay85708 Peptide s
32	28	33.7	16	4	AAB45618	Aab45618 Vasoactiv
33	28	33.7	16	4	AAB45620	Aab45620 Vasoactiv
34	28	33.7	17	2	AAR79546	Aar79546 Analgesic
35	28	33.7	17	4	AAB45617	Aab45617 Vasoactiv
36	28	33.7	17	6	ADA90426	Ada90426 MS-Roche
37	28	33.7	17	6	ADA89996	Ada89996 Anti-Abet
38	27	32.5	8	4	ABP19278	Abp19278 HIV B62 s
39	27	32.5	8	4	ABP19181	Abp19181 HIV B62 s
40	27	32.5	9	4	ABP21094	Abp21094 HIV A03 m
41	27	32.5	9	4	ABP23330	Abp23330 HIV A11 m
42	27	32.5	9	4	ABP21373	Abp21373 HIV A03 m
43	27	32.5	9	4	ABP23124	Abp23124 HIV A11 m
44	27	32.5	9	7	ADD57378	Add57378 HLA bindi
45	27	32.5	9	7	ADD57758	Add57758 HLA bindi

ALIGNMENTS

RESULT 1
AAY97148
ID AAY97148 standard; peptide; 17 AA.
XX
AC AAY97148;
XX
DT 04-DEC-2000 (first entry)
XX
DE Pigment protein from coral tissue N-terminal peptide 2.
XX
KW N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW UV filter.
XX
OS Acropora horrida.
XX
PN WO200046233-A1.
XX
PD 10-AUG-2000.
XX
PF 02-FEB-2000; 2000WO-AU000056.
XX
PR 02-FEB-1999; 99AU-00008463.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Hoegh-Guldberg O, Dove S;
XX
DR WPI; 2000-532892/48.
XX
PT Novel pigment protein derived from corals capable of emitting
PT fluorescence upon irradiation by incident light useful as tissue marker,
PT fluorescent marker or general dyestuff.
XX
PS Claim 4; Page 42; 49pp; English.
XX
CC The N-terminal peptides shown in AAY97147-48 are from pigment protein
CC from coral tissue (PPCT). PPCT is capable of emitting fluorescence upon
CC irradiation by incident light whose maximal absorbance is in the range of
CC 320-600 nm and a maximal fluorescence emission is in the range of 300-700
CC nm. PPCT may be used as a tissue marker, fluorescent marker (e.g. to
CC follow gene expression in transformed tissues) or general dyestuff (all
CC claimed). PPCT may also be used in sunscreen formulations or UV filters
CC (both claimed)
XX
SQ Sequence 17 AA;

Query Match 100.0%; Score 83; DB 3; Length 17;

```
  Best Local Similarity   100.0%;  Pred. No. 3.6e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTV 17


RESULT 2
ABB99066
ID   ABB99066 standard; peptide; 16 AA.
XX
AC   ABB99066;
XX
DT   22-JAN-2003  (first entry)
XX
DE   N-terminal amino acid sequence of a CFM #6.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 4; Page 280; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 16 AA;

  Query Match             88.0%;  Score 73;  DB 5;  Length 16;
  Best Local Similarity   93.8%;  Pred. No. 2.2e-06;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              |||| ||||||||||
Db          1 SVIATQMTYKVYMSGT 16


RESULT 3
ABB99073
ID   ABB99073 standard; peptide; 16 AA.
XX
AC   ABB99073;
XX
DT   22-JAN-2003  (first entry)
XX
DE   N-terminal amino acid sequence of a CFM #13.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 4; Page 281; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
```

```
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 16 AA;

  Query Match             86.7%;  Score 72;  DB 5;  Length 16;
  Best Local Similarity   93.8%;  Pred. No. 3.4e-06;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              ||||||||||||||| |
Db          1 SVIAKQMTYKVYMSDT 16


RESULT 4
ABB99072
ID   ABB99072 standard; peptide; 16 AA.
XX
AC   ABB99072;
XX
DT   22-JAN-2003  (first entry)
XX
DE   N-terminal amino acid sequence of a CFM #12.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 4; Page 281; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 16 AA;

  Query Match             84.3%;  Score 70;  DB 5;  Length 16;
  Best Local Similarity   93.8%;  Pred. No. 7.9e-06;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              |||||||||| ||||
Db          1 SVIAKQMTYKVNMSGT 16


RESULT 5
ABB99068
ID   ABB99068 standard; peptide; 16 AA.
XX
AC   ABB99068;
XX
DT   22-JAN-2003  (first entry)
XX
DE   N-terminal amino acid sequence of a CFM #8.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 4; Page 280; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
```

```
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 16 AA;

  Query Match             83.1%;  Score 69;  DB 5;  Length 16;
  Best Local Similarity   87.5%;  Pred. No. 1.2e-05;
  Matches   14;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              |||| |:|||||||||
Db          1 SVIATQVTYKVYMSGT 16


RESULT 6
ABB99070
ID   ABB99070 standard; peptide; 16 AA.
XX
AC   ABB99070;
XX
DT   22-JAN-2003  (first entry)
XX
DE   N-terminal amino acid sequence of a CFM #10.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 4; Page 281; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 16 AA;

  Query Match             83.1%;  Score 69;  DB 5;  Length 16;
  Best Local Similarity   87.5%;  Pred. No. 1.2e-05;
  Matches   14;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              |||  ||||||||||
Db          1 SVIVTQMTYKVYMSGT 16


RESULT 7
ABB99067
ID   ABB99067 standard; peptide; 16 AA.
XX
AC   ABB99067;
XX
DT   22-JAN-2003  (first entry)
XX
DE   N-terminal amino acid sequence of a CFM #7.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 4; Page 280; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
```

CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC current sequence represents the N-terminal amino acid sequence of a
CC colour-facilitating molecule (CFM)
XX
SQ Sequence 16 AA;

```
  Query Match             81.9%;  Score 68;  DB 5;  Length 16;
  Best Local Similarity   87.5%;  Pred. No. 1.8e-05;
  Matches   14;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              |||| |||||||| ||
Db          1 SVIATQMTYKVYMPGT 16
```

RESULT 8
ABB99071
ID ABB99071 standard; peptide; 16 AA.
XX
AC ABB99071;
XX
DT 22-JAN-2003 (first entry)
XX
DE N-terminal amino acid sequence of a CFM #11.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Unidentified.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 4; Page 281; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC current sequence represents the N-terminal amino acid sequence of a
CC colour-facilitating molecule (CFM)
XX
SQ Sequence 16 AA;

```
  Query Match             80.7%;  Score 67;  DB 5;  Length 16;
  Best Local Similarity   87.5%;  Pred. No. 2.8e-05;
  Matches   14;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              || | |||||||||||
Db          1 SVSATQMTYKVYMSGT 16
```

RESULT 9
ABB99069
ID ABB99069 standard; peptide; 16 AA.
XX
AC ABB99069;
XX
DT 22-JAN-2003 (first entry)
XX
DE N-terminal amino acid sequence of a CFM #9.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Unidentified.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 4; Page 280; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)

```
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 16 AA;

  Query Match             79.5%;  Score 66;  DB 5;  Length 16;
  Best Local Similarity   87.5%;  Pred. No. 4.3e-05;
  Matches   14;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              | || ||||||||||
Db          1 SGIATQMTYKVYMSGT 16


RESULT 10
ABB99074
ID   ABB99074 standard; peptide; 16 AA.
XX
AC   ABB99074;
XX
DT   22-JAN-2003  (first entry)
XX
DE   N-terminal amino acid sequence of a CFM #14.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 10
FT                   /label= Xaa
FT                   /note= "Xaa is any amino acid except Lys"
FT   Misc-difference 11
FT                   /label= Xaa
FT                   /note= "Xaa is any amino acid except Val"
FT   Misc-difference 13
FT                   /label= Xaa
FT                   /note= "Xaa is any amino acid except Met"
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 4; Page 282; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 16 AA;

  Query Match             74.7%;  Score 62;  DB 5;  Length 16;
  Best Local Similarity   81.2%;  Pred. No. 0.00023;
  Matches   13;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGT 16
              |||||||||  | |||
Db          1 SVIAKQMTYXXYXSGT 16


RESULT 11
ABP70008
ID   ABP70008 standard; peptide; 13 AA.
XX
AC   ABP70008;
XX
DT   06-AUG-2003  (revised)
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 184.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Pavona decussata.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
```

XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 473; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 13 AA;

Query Match 43.4%; Score 36; DB 5; Length 13;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMT 8
||||||||
Db 1 SVIAKQMT 8

RESULT 12
AAR77526
ID AAR77526 standard; peptide; 14 AA.
XX
AC AAR77526;
XX
DT 25-MAR-2003 (revised)
DT 12-JUN-1996 (first entry)
XX
DE p45 metalloprotease N-terminal fragment.
XX
KW Metalloprotease; enzyme; MP; p45; fusarium oxysporum; bacillus;
KW thermolysin; casein; Aspergillus oryzae.
XX
OS Fusarium oxysporum.
XX
PN WO9530757-A2.
XX
PD 16-NOV-1995.
XX
PF 03-MAY-1995; 95WO-US005534.
XX
PR 04-MAY-1994; 94US-00238108.
PR 03-MAR-1995; 95US-00398489.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO-NORDISK AS.
XX
PI Shuster JR, Moyer DL, Madden M, Fuglsang C, Branner S;
XX
DR WPI; 1995-404122/51.
XX
PT Fungal metallo:protease converts pro:enzyme to active form - has
PT thermolysin-like activity, useful to cleave pro-sequence of pro:enzyme to
PT generate mature enzyme.
XX
PS Claim 12; Page 36; 62pp; English.
XX
CC AAR77525-R77527 represent the N-terminal sequences of a fungal
CC metalloprotease (MP). This sequence represents the N-terminus of Fusarium
CC oxysporum MP p45 (see AAR77528). AAR77525 represents the consensus N-
CC terminal sequence of the MP from F.oxysporum and Aspergillus oryzae. p45
CC is a new MP, and has 10 times more efficiency than Bacillus MP. Bacillus
CC MP is more effective in cleaving primary amino groups from casein. p45
CC has thermolysin-like activity, and is used to cleave a pro-sequence from
CC a recombinant proenzyme to generate an active mature enzyme. The MP may
CC be added to, or produced in, the broth where the proenzyme is being
CC formed by a recombinant host cell converted with a vector containing the
CC DNA encoding p45. The MP can also be used to assay the level of
CC activatable proenzyme in a sample. (Updated on 25-MAR-2003 to correct PA
CC field.)
XX
SQ Sequence 14 AA;

Query Match 34.9%; Score 29; DB 2; Length 14;
Best Local Similarity 75.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 8 TYKVYMSG 15
|||||
Db 2 TYKVYPWG 9

RESULT 13
AAW05846
ID AAW05846 standard; peptide; 14 AA.
XX
AC AAW05846;
XX
DT 16-OCT-2003 (revised)
DT 28-JAN-1997 (first entry)
XX
DE Fusarium oxysporum p45 metalloprotease N-terminal peptide.
XX
KW Metalloprotease; protease; p45; recombinant protein; host cell.
XX
OS Fusarium oxysporum; strain DSM 2672.
XX
PN WO9629391-A1.
XX
PD 26-SEP-1996.
XX
PF 20-MAR-1996; 96WO-DK000111.
XX
PR 20-MAR-1995; 95DK-00000284.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Lehmbeck J;
XX
DR WPI; 1996-443168/44.
XX
PT Host cell with reduced expression of metallo-protease - for prodn. of
PT recombinant proteins, opt. as their precursors.
XX
PS Example 1; Page 34; 51pp; English.

```
XX
CC   The N-terminal sequence (AAW05846) of Fusarium oxysporum DSM 2672 p45
CC   metalloprotease (see also AAW05845) was identified by amino acid analysis
CC   of a protein isolated from a fermentation broth. A PCR primer based on
CC   this peptide was used, together with a primer based on a p45 internal
CC   peptide, in the PCR cloning of the p45 gene (AAT40133) from F. oxysporum
CC   genomic DNA. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 14 AA;

  Query Match             34.9%;  Score 29;  DB 2;  Length 14;
  Best Local Similarity   75.0%;  Pred. No. 2.2e+02;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          8 TYKVYMSG 15
              |||||  |
Db          2 TYKVYPWG 9


RESULT 14
AAM48968
ID   AAM48968 standard; peptide; 15 AA.
XX
AC   AAM48968;
XX
DT   25-APR-2002  (first entry)
XX
DE   Human zinc finger protein 53 N-terminal peptide.
XX
KW   Human; zinc finger protein 53; cancer; nervous system disease;
KW   development disorder; metabolic disease; inflammation; haemopathy;
KW   immunological disease; HIV infection; gene therapy.
XX
OS   Homo sapiens.
XX
PN   CN1314368-A.
XX
PD   26-SEP-2001.
XX
PF   17-MAR-2000; 2000CN-00114979.
XX
PR   17-MAR-2000; 2000CN-00114979.
XX
PA   (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX
PI   Mao Y,  Xie Y;
XX
DR   WPI; 2002-056224/08.
XX
PT   New polypeptide-human zinc finger protein 53 and polynucleotide for
PT   coding such polypeptide.
XX
PS   Example 6; Page 18(Disclosure); 33pp; Chinese.
XX
CC   The present invention provides the protein and coding sequences of human
CC   zinc finger protein 53. The sequences can be used in the treatment of
CC   cancer, haemopathy, nervous system disorders, development disorders,
CC   metabolic disorders, inflammation, immunological diseases and HIV
CC   infection. The present sequence is the N-terminus of the protein of the
CC   invention
XX
SQ   Sequence 15 AA;

  Query Match             34.9%;  Score 29;  DB 5;  Length 15;
  Best Local Similarity   54.5%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          5 KQMTYKVYMSG 15
              | || | : ||
Db          2 KNMTLKSFASG 12


RESULT 15
AAW39598
ID   AAW39598 standard; peptide; 11 AA.
XX
AC   AAW39598;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human melanoma associated protein tyrosinase peptide (pos. 367-377).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
PF   28-APR-1997;   97WO-NL000229.
XX
PR   26-APR-1996;   96EP-00201145.
PR   23-DEC-1996;   96EP-00203670.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI   Van Der Burg SH,  Kast WM,  Toes REM,  Offringa R,  Melief CJM;
XX
DR   WPI; 1997-549891/50.
XX
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells.
XX
PS   Example 3; Page 75; 109pp; English.
XX
CC   Peptides AAW39430-W39734 are used in a novel method for the selection of
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC   method involves the identification of peptide sequences capable of
CC   binding to an HLA (human leukocyte antigen) class I molecule and
CC   measuring the binding of this epitope peptide to the HLA class I peptide.
CC   The stability of binding of the peptide and MHC (major histocompatablilty
CC   complex) class I molecule is measured on intact human B cells carrying
CC   the MHC molecule at their cell surfaces. The method can be used to select
CC   peptide epitopes for generating vaccines against a disease associated
CC   with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC   especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC   immune responses. Peptide AAW39598 is derived from the human melanoma
CC   associated protein tyrosinase which is capable of upregulating HLA-A*0201
CC   molecules on T2 cells
XX
SQ   Sequence 11 AA;

  Query Match             33.7%;  Score 28;  DB 2;  Length 11;
  Best Local Similarity   57.1%;  Pred. No. 2.5e+02;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          11 VYMSGTV 17
                :||:||:
Db           2 IYMNGTM 8


Search completed: August 12, 2004, 07:03:22
Job time : 51 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 89.0881 Seconds
(without alignments)
745.314 Million cell updates/sec

Title: US-09-890-463-4
Perfect score: 1287
Sequence: 1 SVIAKQMTYKVYMSGTVNGH.....KPVVACRFFRVKSRHKYAVA 235

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1287	100.0	235	3	AAY97150	Aay97150 Pigment p
2	1279	99.4	235	5	ABP70042	Abp70042 Colour Fa
3	1257	97.7	235	5	ABP70026	Abp70026 Colour Fa
4	1242	96.5	231	3	AAY97149	Aay97149 Pigment p
5	1242	96.5	231	5	ABP70025	Abp70025 Colour Fa
6	1190	92.5	226	5	ABP70036	Abp70036 Colour Fa
7	1188	92.3	220	5	ABP70037	Abp70037 Colour Fa
8	1174	91.2	220	5	ABP69926	Abp69926 Colour Fa
9	1174	91.2	223	5	ABP70030	Abp70030 Colour Fa
10	1174	91.2	223	5	ABP70032	Abp70032 Colour Fa
11	1169	90.8	221	5	ABP69992	Abp69992 Colour Fa
12	1169	90.8	221	5	ABP69991	Abp69991 Colour Fa
13	1166	90.6	220	5	ABP70007	Abp70007 Colour Fa
14	1165	90.5	221	5	ABP69967	Abp69967 Colour Fa
15	1165	90.5	221	5	ABP69966	Abp69966 Colour Fa
16	1165	90.5	221	5	ABP70004	Abp70004 Colour Fa
17	1165	90.5	223	5	ABP70033	Abp70033 Colour Fa
18	1165	90.5	235	5	ABP69963	Abp69963 Colour Fa
19	1165	90.5	235	5	ABP69961	Abp69961 Colour Fa
20	1164	90.4	221	5	ABP69978	Abp69978 Colour Fa
21	1163	90.4	223	5	ABP70029	Abp70029 Colour Fa
22	1162	90.3	220	5	ABP69941	Abp69941 Colour Fa
23	1162	90.3	220	5	ABP69940	Abp69940 Colour Fa
24	1161	90.2	220	5	ABP69952	Abp69952 Colour Fa
25	1161	90.2	220	5	ABP69959	Abp69959 Colour Fa
26	1161	90.2	220	5	ABP69934	Abp69934 Colour Fa
27	1161	90.2	220	5	ABP69958	Abp69958 Colour Fa
28	1161	90.2	220	5	ABP69937	Abp69937 Colour Fa
29	1161	90.2	220	5	ABP69935	Abp69935 Colour Fa
30	1161	90.2	221	5	ABP69972	Abp69972 Colour Fa
31	1159.5	90.1	222	5	ABP70027	Abp70027 Colour Fa
32	1159	90.1	220	5	ABP69939	Abp69939 Colour Fa
33	1158	90.0	220	5	ABP69965	Abp69965 Colour Fa
34	1158	90.0	221	5	ABP70002	Abp70002 Colour Fa
35	1157	89.9	220	5	ABP69938	Abp69938 Colour Fa
36	1157	89.9	223	5	ABP70031	Abp70031 Colour Fa
37	1156	89.8	220	5	ABP69925	Abp69925 Colour Fa
38	1156	89.8	220	5	ABP69964	Abp69964 Colour Fa
39	1155	89.7	220	5	ABP69993	Abp69993 Colour Fa
40	1155	89.7	220	5	ABP69989	Abp69989 Colour Fa
41	1155	89.7	221	5	ABP69986	Abp69986 Colour Fa
42	1154	89.7	220	5	ABP69930	Abp69930 Colour Fa
43	1152	89.5	220	5	ABP69999	Abp69999 Colour Fa
44	1152	89.5	221	5	ABP70035	Abp70035 Colour Fa
45	1151	89.4	220	5	ABP69936	Abp69936 Colour Fa

ALIGNMENTS

RESULT 1
AAY97150
ID AAY97150 standard; protein; 235 AA.
XX
AC AAY97150;
XX
DT 04-DEC-2000 (first entry)
XX
DE Pigment protein from coral tissue POC4.
XX
KW N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW UV filter; POC3.
XX
OS Acropora aspera.
XX
FH Key Location/Qualifiers
FT Misc-difference 61. .63
FT /label= Chromophore_motif
FT Misc-difference 158
FT /note= "critical residue in the vicinity of the
FT fluorophore"
FT Misc-difference 192
FT /note= "critical residue in the vicinity of the
FT fluorophore"
FT Misc-difference 210
FT /note= "critical residue in the vicinity of the
FT fluorophore"
XX
PN WO200046233-A1.
XX
PD 10-AUG-2000.
XX
PF 02-FEB-2000; 2000WO-AU000056.
XX
PR 02-FEB-1999; 99AU-00008463.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Hoegh-Guldberg O, Dove S;
XX
DR WPI; 2000-532892/48.
DR N-PSDB; AAA52083.
XX
PT Novel pigment protein derived from corals capable of emitting
PT fluorescence upon irradiation by incident light useful as tissue marker,
PT fluorescent marker or general dyestuff.
XX

PS Claim 13; Page 43-44; 49pp; English.
XX
CC cDNA libraries were constructed from a blue pigmented coral, Acropora
CC aspera to isolate sequences encoding polypeptides with N-terminal
CC sequences as shown in AAY97147-48. Pigment protein from coral tissue
CC (PPCT) is capable of emitting fluorescence upon irradiation by incident
CC light whose maximal absorbance is in the range of 320-600 nm and a
CC maximal fluorescence emission is in the range of 300-700 nm. PPCT may be
CC used as a tissue marker, fluorescent marker (e.g. to follow gene
CC expression in transformed tissues) or general dyestuff (all claimed).
CC PPCT may also be used in sunscreen formulations or UV filters (both
CC claimed)
XX
SQ Sequence 235 AA;

Query Match 100.0%; Score 1287; DB 3; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e-127;
Matches 235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
Db 1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60

Qy 61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
Db 61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy 121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
Db 121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180

Qy 181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVACRFFRVKSRHKYAVA 235
Db 181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVACRFFRVKSRHKYAVA 235

RESULT 2
ABP70042
ID ABP70042 standard; protein; 235 AA.
XX
AC ABP70042;
XX
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 245.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Acropora aspera.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Example 20; Page 502-503; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 235 AA;

Query Match 99.4%; Score 1279; DB 5; Length 235;
Best Local Similarity 99.6%; Pred. No. 8.7e-127;
Matches 234; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
Db 1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60

Qy 61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
Db 61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy 121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
Db 121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180

Qy 181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVACRFFRVKSRHKYAVA 235
Db 181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVACRFFRVKSRHKVAVA 235

RESULT 3
ABP70026
ID ABP70026 standard; protein; 235 AA.
XX
AC ABP70026;
XX
DT 06-AUG-2003 (revised)
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 202.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Unidentified.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX

PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 479; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 235 AA;

Query Match 97.7%; Score 1257; DB 5; Length 235;
Best Local Similarity 98.3%; Pred. No. 1.9e-124;
Matches 231; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
Db 1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGLPYEGGQTVRLAVTKGGPLPFAWDILSPQC 60

Qy 61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
Db 61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKRS 120

Qy 121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
Db 121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180

Qy 181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVACRFFRVKSRHKYAVA 235
Db 181 PVKMPGYHYVDRKLDVTNHNLDYTSVEQCEISIARKPVVACRFFRVKSRHKYAVA 235

RESULT 4
AAY97149
ID AAY97149 standard; protein; 231 AA.
XX
AC AAY97149;
XX
DT 04-DEC-2000 (first entry)
XX
DE Pigment protein from coral tissue POC3.
XX
KW N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW UV filter; POC3.
XX
OS Acropora aspera.
XX
FH Key Location/Qualifiers
FT Misc-difference 61. .63
FT /label= Chromophore_motif
FT Misc-difference 158
FT /note= "critical residue in the vicinity of the
FT fluorophore"
FT Misc-difference 192
FT /note= "critical residue in the vicinity of the
FT fluorophore"
FT Misc-difference 210
FT /note= "critical residue in the vicinity of the
FT fluorophore"
XX
PN WO200046233-A1.
XX
PD 10-AUG-2000.
XX
PF 02-FEB-2000; 2000WO-AU000056.
XX
PR 02-FEB-1999; 99AU-00008463.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Hoegh-Guldberg O, Dove S;
XX
DR WPI; 2000-532892/48.
DR N-PSDB; AAA52082.
XX
PT Novel pigment protein derived from corals capable of emitting
PT fluorescence upon irradiation by incident light useful as tissue marker,
PT fluorescent marker or general dyestuff.
XX
PS Claim 13; Page 42-43; 49pp; English.
XX
CC cDNA libraries were constructed from a blue pigmented coral, Acropora
CC aspera to isolate sequences encoding polypeptides with N-terminal
CC sequences as shown in AAY97147-48. Pigment protein from coral tissue
CC (PPCT) is capable of emitting fluorescence upon irradiation by incident
CC light whose maximal absorbance is in the range of 320-600 nm and a
CC maximal fluorescence emission is in the range of 300-700 nm. PPCT may be
CC used as a tissue marker, fluorescent marker (e.g. to follow gene
CC expression in transformed tissues) or general dyestuff (all claimed).
CC PPCT may also be used in sunscreen formulations or UV filters (both
CC claimed)
XX
SQ Sequence 231 AA;

Query Match 96.5%; Score 1242; DB 3; Length 231;
Best Local Similarity 98.3%; Pred. No. 7e-123;
Matches 227; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
Db 1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60

Qy 61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
Db 61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy 121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
Db 121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180

```
Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVACRFFRVKSRHK 231
              ||||||||||||||||||||||||||||| |||||||||:||| |||||||||
Db        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVACCFFRVKSRHK 231
```

RESULT 5
ABP70025
ID ABP70025 standard; protein; 231 AA.
XX
AC ABP70025;
XX
DT 06-AUG-2003 (revised)
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 201.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Unidentified.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 6; Page 478; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 231 AA;

Query Match 96.5%; Score 1242; DB 5; Length 231;
Best Local Similarity 98.3%; Pred. No. 7e-123;
Matches 227; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVACRFFRVKSRHK 231
              ||||||||||||||||||||||||||| |||||||:||| |||||||||
Db        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVACCFFRVKSRHK 231
```

RESULT 6
ABP70036
ID ABP70036 standard; protein; 226 AA.
XX
AC ABP70036;
XX
DT 06-AUG-2003 (revised)
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 238.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Unidentified.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Example 19; Page 496-497; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other

```
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 226 AA;

  Query Match             92.5%;  Score 1190;  DB 5;  Length 226;
  Best Local Similarity   96.9%;  Pred. No. 2.2e-117;
  Matches  219;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLTVTKGGPLPFAWDILSPQS 60

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db         61 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVACRFFRV 226
              ||:|||||||||||||||||||||||||||||||||||  ||||
Db        181 PVRMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVAWCFFRV 226


RESULT 7
ABP70037
ID   ABP70037 standard; protein; 220 AA.
XX
AC   ABP70037;
XX
DT   06-AUG-2003  (revised)
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 239.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Example 19; Page 497-498; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 220 AA;

  Query Match             92.3%;  Score 1188;  DB 5;  Length 220;
  Best Local Similarity   98.6%;  Pred. No. 3.4e-117;
  Matches  217;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 220
              |||||||||||||||||||||||||||| ||||||||:||
Db        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVA 220


RESULT 8
ABP69926
ID   ABP69926 standard; protein; 220 AA.
XX
AC   ABP69926;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 24.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Acropora aspera.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
```

```
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 289; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences
XX
SQ   Sequence 220 AA;

  Query Match             91.2%;  Score 1174;  DB 5;  Length 220;
  Best Local Similarity   97.7%;  Pred. No. 1e-115;
  Matches  215;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLTVTKGGPLPFAWDILSPQS 60

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db         61 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 220
              ||:|||||||||||||||||||||||||||||||||||||
Db        181 PVRMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 220


RESULT 9
ABP70030
ID   ABP70030 standard; protein; 223 AA.
XX
AC   ABP70030;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 216.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Tubastrea sp.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Example 18; Page 486; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences
XX
SQ   Sequence 223 AA;

  Query Match             91.2%;  Score 1174;  DB 5;  Length 223;
  Best Local Similarity   97.7%;  Pred. No. 1.1e-115;
  Matches  215;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||||||||||||||||||||||||||||||||||||||:| ||||||||||||||||||
Db          2 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVKLTVTKGGPLPFAWDILSPQS 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db         62 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 121
```

```
Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 181

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 220
              |||||||||||||||||||||||||||||||||||||||||
Db        182 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 221
```

RESULT 10
ABP70032
ID ABP70032 standard; protein; 223 AA.
XX
AC ABP70032;
XX
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 220.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Sinularia sp.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Disclosure; Page 489; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 223 AA;

```
  Query Match             91.2%;  Score 1174;  DB 5;  Length 223;
  Best Local Similarity   97.7%;  Pred. No. 1.1e-115;
  Matches  215;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||| |||||||||||  :|||||||||||||||||||||||||||||||||||||
Db         62 QYGSIPFTKYLEDIPDYVKQSFPEGFTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 121

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 181

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 220
              |||||||||||||||||||||||||||||||||||||:||
Db        182 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPLVA 221
```

RESULT 11
ABP69992
ID ABP69992 standard; protein; 221 AA.
XX
AC ABP69992;
XX
DT 06-AUG-2003 (revised)
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 149.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Pocillopora sp.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 6; Page 435-436; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous

```
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 221 AA;

  Query Match             90.8%;  Score 1169;  DB 5;  Length 221;
  Best Local Similarity   96.8%;  Pred. No. 3.6e-115;
  Matches  213;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 SVIATQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||||||||||||||||  :|||||||||||||||||||||||||||||||||
Db         62 QYGSIPFTKYPEDIPDYVKQSFPEGFTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 121

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              ||||||||||||||||||::|||||||||||||||||||||||||:||||||
Db        122 GLNFPPNGPVMQKKTQGWEPHSERLFARDGMLIGNNFMALKLEGGGHYLCEFKTTYKAKK 181

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 220
              |||||||||||||||||||||||||||||||||||||||
Db        182 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 221


RESULT 12
ABP69991
ID   ABP69991 standard; protein; 221 AA.
XX
AC   ABP69991;
XX
DT   06-AUG-2003  (revised)
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 147.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Pocillopora sp.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 6; Page 433-434; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 221 AA;

  Query Match             90.8%;  Score 1169;  DB 5;  Length 221;
  Best Local Similarity   96.8%;  Pred. No. 3.6e-115;
  Matches  213;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 SVIATQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||||||||||||||||  :|||||||||||||||||||||||||||||||||
Db         62 QYGSIPFTKYPEDIPDYVKQSFPEGFTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 121

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              ||||||||||||||||||::|||||||||||||||||||||||||:||||||
Db        122 GLNFPPNGPVMQKKTQGWEPHSERLFARDGMLIGNNFMALKLEGGGHYLCEFKTTYKAKK 181

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 220
              |||||||||||||||||||||||||||||||||||||||
Db        182 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 221


RESULT 13
ABP70007
ID   ABP70007 standard; protein; 220 AA.
XX
AC   ABP70007;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 177.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Montipora sp.
```

```
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 6; Page 467-468; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences
XX
SQ   Sequence 220 AA;

  Query Match             90.6%;  Score 1166;  DB 5;  Length 220;
  Best Local Similarity   97.3%;  Pred. No. 7.4e-115;
  Matches  214;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||  |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db          1 SVIVTQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLTVTKGGPLPFAWDILSPQY 60

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db         61 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 220
              |||||||||||||||||||||||||||||||||||||||||
Db        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 220


RESULT 14
ABP69967
ID   ABP69967 standard; protein; 221 AA.
XX
AC   ABP69967;
XX
DT   06-AUG-2003  (revised)
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 102.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Acropora aspera.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 383-384; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 221 AA;

  Query Match             90.5%;  Score 1165;  DB 5;  Length 221;
  Best Local Similarity   96.4%;  Pred. No. 9.5e-115;
  Matches  212;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          2 SVIATQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||||||||||||||||  :|||||||||||||||||||||||||||||||||||
Db         62 QYGSIPFTKYPEDIPDYVKQSFPEGFTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 121

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              |||||||||||||||||||::|||||||||||||||||||||||||||||:||||||
Db        122 GLNFPPNGPVMQKKTQGWEPHSERLFARDGMLIGNNFMALKLEGGGHYLCEFKTTYKAKK 181

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 220
              |||||||||||||||||||||||||||||| ||||||
Db        182 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISITRKPVVA 221


RESULT 15
ABP69966
ID   ABP69966 standard; protein; 221 AA.
XX
AC   ABP69966;
XX
DT   06-AUG-2003  (revised)
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 100.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Acropora aspera.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 381-382; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 221 AA;

  Query Match             90.5%;  Score 1165;  DB 5;  Length 221;
  Best Local Similarity   96.4%;  Pred. No. 9.5e-115;
  Matches  212;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 SVIATQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||||||||||||||  :|||||||||||||||||||||||||||||||||||||
Db         62 QYGSIPFTKYPEDIPDYVKQSFPEGFTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 121

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              |||||||||||||||||||::|||||||||||||||||||||||||||||:||||||
Db        122 GLNFPPNGPVMQKKTQGWEPHSERLFARDGMLIGNNFMALKLEGGGHYLCEFKTTYKAKK 181

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 220
              |||||||||||||||||||||||||||||| ||||||
Db        182 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISITRKPVVA 221


Search completed: August 12, 2004, 06:17:06
Job time : 90.0881 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 61.1578 Seconds
(without alignments)
1212.385 Million cell updates/sec

Title: US-09-890-463-4
Perfect score: 1287
Sequence: 1 SVIAKQMTYKVYMSGTVNGH.....KPVVACRFFRVKSRHKYAVA 235

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1156	89.8	221	5	Q95P04	Q95p04 goniopora t
2	783	60.8	225	5	Q9U6Y8	Q9u6y8 discosoma s
3	750	58.3	230	5	Q9GTJ7	Q9gtj7 discosoma s
4	737.5	57.3	232	5	Q9U6Y7	Q9u6y7 discosoma s
5	707	54.9	236	5	Q8T6U0	Q8t6u0 dendronepht
6	706	54.9	225	5	Q963F5	Q963f5 montastraea
7	701	54.5	225	5	Q7Z0W4	Q7z0w4 montastraea
8	674	52.4	225	5	Q7Z0W5	Q7z0w5 montastraea
9	671	52.1	266	5	Q9U6Y3	Q9u6y3 clavularia
10	669.5	52.0	227	5	Q7Z0W6	Q7z0w6 montastraea
11	667.5	51.9	225	5	Q7Z0W9	Q7z0w9 montastraea
12	667.5	51.9	227	5	Q962P9	Q962p9 montastraea
13	667.5	51.9	227	5	Q7Z0W8	Q7z0w8 montastraea
14	666.5	51.8	225	5	Q95UA7	Q95ua7 montastraea
15	656.5	51.0	225	5	Q8T5F1	Q8t5f1 montastraea
16	654.5	50.9	227	5	Q95VT0	Q95vt0 montastraea
17	654.5	50.9	234	5	Q8T5F2	Q8t5f2 montastraea
18	654	50.8	224	5	Q8MU48	Q8mu48 montastraea
19	653.5	50.8	234	5	Q7Z0W7	Q7z0w7 montastraea
20	631.5	49.1	225	5	Q8I6J8	Q8i6j8 trachyphyll
21	625.5	48.6	234	5	Q8MU47	Q8mu47 montastraea
22	612.5	47.6	259	5	Q8MMA2	Q8mma2 agaricia fr
23	609.5	47.4	231	5	Q8T5E9	Q8t5e9 ricordea fl
24	601	46.7	231	5	Q8T5E8	Q8t5e8 ricordea fl
25	600.5	46.7	231	5	Q8T6T8	Q8t6t8 discosoma s
26	586.5	45.6	231	5	Q8ISF8	Q8isf8 parasicyoni
27	579.5	45.0	239	5	Q8MMA1	Q8mma1 agaricia ag
28	578.5	44.9	231	5	Q8MU46	Q8mu46 ricordea fl
29	574.5	44.6	232	5	Q9GPI5	Q9gpi5 anemonia su
30	569.5	44.3	227	5	Q95W86	Q95w86 condylactis
31	565.5	43.9	227	5	Q95W85	Q95w85 radianthus
32	561.5	43.6	227	5	Q95W11	Q95w11 condylactis
33	561.5	43.6	232	5	Q9GZ28	Q9gz28 anemonia su
34	561	43.6	228	5	Q9GPI6	Q9gpi6 anemonia su
35	549.5	42.7	227	5	Q8MU45	Q8mu45 condylactis
36	536.5	41.7	235	5	Q8T5F0	Q8t5f0 scolymia cu
37	534	41.5	214	5	Q86LV7	Q86lv7 meandrina m
38	534	41.5	228	5	Q86LV4	Q86lv4 radianthus
39	531	41.3	229	5	Q9U6Y6	Q9u6y6 anemonia ma
40	529.5	41.1	225	5	Q8T6T9	Q8t6t9 radianthus
41	529	41.1	214	5	Q86LV8	Q86lv8 meandrina m
42	527.5	41.0	234	5	Q8T5F3	Q8t5f3 scolymia cu
43	516.5	40.1	229	5	Q8T5E7	Q8t5e7 condylactis
44	488.5	38.0	231	5	Q9U6Y5	Q9u6y5 zoanthus sp
45	480.5	37.3	231	5	Q9U6Y4	Q9u6y4 zoanthus sp

ALIGNMENTS

RESULT 1
Q95P04
ID Q95P04 PRELIMINARY; PRT; 221 AA.
AC Q95P04;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GFP-like chromoprotein.
OS Goniopora tenuidens.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Fungiina; Poritidae; Goniopora.
OX NCBI_TaxID=75301;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21538626; PubMed=11682051;
RA Gurskaya N.G., Fradkov A.F., Terskikh A., Matz M.V., Labas Y.A.,
RA Martynov V.I., Yanushevich Y.G., Lukyanov K.A., Lukyanov S.A.;
RT "GFP-like chromoproteins as a source of far-red fluorescent
RT proteins(1).";
RL FEBS Lett. 507:16-20(2001).
DR EMBL; AF383156; AAL27542.1; -.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 221 AA; 24918 MW; 93F9F4B5C2003CB4 CRC64;

Query Match 89.8%; Score 1156; DB 5; Length 221;
Best Local Similarity 96.4%; Pred. No. 2.1e-100;
Matches 212; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||||||||||||||||||||||:|||||||||||||:| ||||||||||||||||||||
Db          2 SVIAKQMTYKVYMSGTVNGHYFEVQGDGKGKPYEGEQTVKLTVTKGGPLPFAWDILSPQS 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||||||||||||||  ||||||||||||||||||||||||||:||||
Db         62 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYNVKFS 121
```

```
Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 181

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 220
              |||||||||||||||||| ||||||||||||||||||||
Db        182 PVKMPGYHYVDRKLDVTNHNIDYTSVEQCEISIARKPVVA 221


RESULT 2
Q9U6Y8
ID   Q9U6Y8      PRELIMINARY;      PRT;   225 AA.
AC   Q9U6Y8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Fluorescent protein FP583.
OS   Discosoma sp.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
OX   NCBI_TaxID=86600;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99436614; PubMed=10504696;
RA   Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
RA   Markelov M.L., Lukyanov S.A.;
RT   "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL   Nat. Biotechnol. 17:969-973(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
RA   Markelov M.L., Lukyanov S.A.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF168419; AAF03369.1; -.
DR   PDB; 1G7K; 07-NOV-01.
DR   PDB; 1GGX; 06-DEC-00.
DR   GO; GO:0006091; P:energy pathways; IEA.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   225 AA;  25931 MW;  FBF9A5369778F689 CRC64;

  Query Match             60.8%;  Score 783;  DB 5;  Length 225;
  Best Local Similarity   63.0%;  Pred. No. 2.2e-65;
  Matches  145;  Conservative   27;  Mismatches   44;  Indels   14;  Gaps    1;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              :|| : | :|| | |||||| ||:|:|:|:|:|||| ||:| ||||||||||||||||||
Db          6 NVIKEFMRFKVRMEGTVNGHFEIEGEGEGRPYEGHNTVKLKVTKGGPLPFAWDILSPQF 65

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              ||||  : |:| ||||| | |||  : |||:|||||| | ||: ||||:  |||| |||
Db         66 QYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVKFI 125

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              |:||| :||||||| ||| :|||: |||:| |    ||||: ||||| |||| | |||
Db        126 GVNFPSDGPVMQKKTMGWEASTERLYPRDGVLKGEIHKALKLKDGGHYLVEFKSIYMAKK 185

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVACRFFRVKSRH 230
              ||::|||:||| |||:|:||:||| ||| |               | : ||
Db        186 PVQLPGYYYVDSKLDITSHNEDYTIVEQYE--------------RTEGRH 221


RESULT 3
Q9GTJ7
ID   Q9GTJ7      PRELIMINARY;      PRT;   230 AA.
AC   Q9GTJ7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Red fluorescent protein.
GN   FP593.
OS   Discosoma sp. SSAL-2000.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
OX   NCBI_TaxID=137428;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20434599; PubMed=10981720;
RA   Fradkov A.F., Chen Y., Ding L., Barsova E.V., Matz M.V.,
RA   Lukyanov S.A.;
RT   "Novel fluorescent protein from Discosoma coral and its mutants
RT   possesses a unique far-red fluorescence.";
RL   FEBS Lett. 479:127-130(2000).
DR   EMBL; AF272711; AAG16224.1; -.
DR   HSSP; P42212; 1BFP.
DR   GO; GO:0006091; P:energy pathways; IEA.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   230 AA;  26370 MW;  5215B1B436D67E51 CRC64;

  Query Match             58.3%;  Score 750;  DB 5;  Length 230;
  Best Local Similarity   62.5%;  Pred. No. 2.8e-62;
  Matches  140;  Conservative   36;  Mismatches   42;  Indels    6;  Gaps    2;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              :|| : | :|| | |||||| ||::|:|:|:|||| :|:| ||||||||||:||||||
Db          6 NVIKEFMRFKVRMEGTVNGHEFEIKGEGEGRPYEGHCSVKLMVTKGGPLPFAFDILSPQF 65

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              ||||  : |:| ||||| | |||  : |||:|||||| | ||| |||::  |||| |||
Db         66 QYGSKVYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGVVTVSQDSSLKDGCFIYEVKFI 125

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              |:||| :|||||::|:||| ::|||: |||:| |:  |||:|||||||| |||| |  ||
Db        126 GVNFPSDGPVMQRRTRGWEASSERLYPRDGVLKGDIHMALRLEGGGHYLVEFKSIYMVKK 185

Qy        181 P-VKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIAR-----KPV 218
              | |::|||:||| |||:|:||:||| ||| | :  |      ||:
Db        186 PSVQLPGYYYVDSKLDMTSHNEDYTVVEQYEKTQGRHHPFIKPL 229


RESULT 4
Q9U6Y7
ID   Q9U6Y7      PRELIMINARY;      PRT;   232 AA.
AC   Q9U6Y7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Fluorescent protein FP483.
OS   Discosoma striata.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC   Discosomatidae; Discosoma.
OX   NCBI_TaxID=105400;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99436614; PubMed=10504696;
RA   Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
RA   Markelov M.L., Lukyanov S.A.;
RT   "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL   Nat. Biotechnol. 17:969-973(1999).
DR   EMBL; AF168420; AAF03370.1; -.
DR   GO; GO:0006091; P:energy pathways; IEA.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   232 AA;  26435 MW;  AA8F18EEE283CE4D CRC64;
```

```
  Query Match           57.3%;  Score 737.5;  DB 5;  Length 232;
  Best Local Similarity   60.2%;  Pred. No. 4.2e-61;
  Matches  130;  Conservative   38;  Mismatches   47;  Indels    1;  Gaps    1;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              ||| ::|   ::: || ||||||::| |||:| ||  || | ||||||||| | || ||
Db          6 SVIKEEMLIDLHLEGTFNGHYFEIKGKGKGQPNEGTNTVTLEVTKGGPLPFGWHILCPQF 65

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||:  |  :|::| ||:| |||  ||||| |:|||| :| ::|| |: |||| | :||:
Db         66 QYGNKAFVHHPDNIHDYLKLSFPEGYTWERSMHFEDGGLCCITNDISLTGNCFYYDIKFT 125

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              |||||||||:|||| ||||:|||: |||:|||:   || :|||||  |: |: |:|||
Db        126 GLNFPPNGPVVQKKTTGWEPSTERLYPRDGVLIGDIHHALTVEGGGHYACDIKTVYRAKK 185

Qy        181 -PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIAR 215
               :|||||||| || : |::|::  ||: ||::||
Db        186 AALKMPGYHYVDTKLVIWNNDKEFMKVEEHEIAVAR 221


RESULT 5
Q8T6U0
ID   Q8T6U0      PRELIMINARY;      PRT;   236 AA.
AC   Q8T6U0;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Green fluorescent protein.
OS   Dendronephthya sp. SSAL-2002.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Alcyonaria; Alcyonacea;
OC   Nephtheidae; Dendronephthya.
OX   NCBI_TaxID=191210;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21927629; PubMed=11929996;
RA   Labas Y.A., Gurskaya N.G., Yanushevich Y.G., Fradkov A.F.,
RA   Lukyanov K.A., Lukyanov S.A., Matz M.V.;
RT   "Diversity and evolution of the green fluorescent protein family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:4256-4261(2002).
DR   EMBL; AF420591; AAM10625.1; -.
DR   GO; GO:0006091; P:energy pathways; IEA.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   236 AA;  26840 MW;  CE1707CFF9334A90 CRC64;

  Query Match           54.9%;  Score 707;  DB 5;  Length 236;
  Best Local Similarity   55.8%;  Pred. No. 3.1e-58;
  Matches  120;  Conservative   45;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              ::| : |  ||:| | |||| | :||:|||:|||| ||: | | :| ||||::|||:
Db          2 NLIKEDMRVKVHMEGNVNGHAFVIEGEGKGRPYEGTQTLNLTVKEGAPLPFSYDILTTAL 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
                 ||:  ||:|| || || |||||  |:||| | :||  :||: :| |::|:||  :::|:
Db         62 HYGNRVFTEYPADITDYFKQSFPEGYSWERTMTYEDKGICTIRSDISLEGDCFFQNIRFN 121

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              |:|||||||||||||  |||:||:|  |||:|:||  ||| ||||||||:||:||||||
Db        122 GMNFPPNGPVMQKKTLKWEPSTEKLHVRDGLLVGNINMALLLEGGGHYLCDFKTTYKAKK 181

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIAR 215
               |::| ||:|| :::: ::: ||  |:  |  :||
Db        182 VVQLPDYHFVDHRIEILSNDSDYNKVKLYEHGVAR 216


RESULT 6
Q963F5
ID   Q963F5      PRELIMINARY;      PRT;   225 AA.
AC   Q963F5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Green fluorescent protein.
OS   Montastraea cavernosa (great star coral).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC   Faviina; Faviidae; Montastraea.
OX   NCBI_TaxID=63558;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lesser M.P., Barry T.M., Mazel C., Matz M.V., Lukyanov S.A.,
RA   Falkowski P., Gorbunov M., Kolber Z.;
RT   "Green fluorescent proteins in Caribbean Scleractinian corals.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF384683; AAK62982.2; -.
DR   GO; GO:0006091; P:energy pathways; IEA.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   225 AA;  25847 MW;  77DE7D7C616929AF CRC64;

  Query Match           54.9%;  Score 706;  DB 5;  Length 225;
  Best Local Similarity   56.7%;  Pred. No. 3.6e-58;
  Matches  122;  Conservative   45;  Mismatches   48;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||   |  |: | | |||| | ::|:|:|||:|| ||: | | :| |||||:|||:
Db          2 SVIKPIMEIKLRMQGVVNGHKFVIKGEGEGKPFEGTQTINLTVKEGAPLPFAYDILTSAF 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||:  |||||:||||| ||:||  |:|||||  :|| ::|| ::|  ::|:||||  ::|
Db         62 QYGNRVFTKYPDDIPDYFKQTFPEGYSWERIMAYEDQSICTATSDIKMEGDCFIYEIQFH 121

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              |:|||||||||||||  |||:|||:::  |||:|  |:  ||| |||||||  |:|:|||||||
Db        122 GVNFPPNGPVMQKKTLKWEPSTEKMYVRDGVLKGDVNMALLLEGGGHYRCDFRSTYKAKK 181

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIAR 215
               |::| ||:|| :::: :|: || :|:  | : ||
Db        182 RVQLPDYHFVDHRIEILSHDNDYNTVKLSEDAEAR 216


RESULT 7
Q7Z0W4
ID   Q7Z0W4      PRELIMINARY;      PRT;   225 AA.
AC   Q7Z0W4;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Green fluorescent protein.
OS   Montastraea cavernosa (great star coral).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC   Faviina; Faviidae; Montastraea.
OX   NCBI_TaxID=63558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=mc6;
RX   MEDLINE=22689801; PubMed=12777529;
RA   Kelmanson I.V., Matz M.V.;
RT   "Molecular Basis and Evolutionary Origins of Color Diversity in Great
RT   Star Coral Montastraea cavernosa (Scleractinia: Faviida).";
RL   Mol. Biol. Evol. 20:1125-1133(2003).
DR   EMBL; AY181557; AAO61603.1; -.
SQ   SEQUENCE   225 AA;  25827 MW;  A600ADD716C5921E CRC64;

  Query Match           54.5%;  Score 701;  DB 5;  Length 225;
  Best Local Similarity   56.3%;  Pred. No. 1.1e-57;
  Matches  121;  Conservative   42;  Mismatches   52;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
```

```
Db          2 SVIKPDMKIKLRMEGAVNGHNFVIEGEGKGKPFEGTQTINLTVKEGGPLPFAYDILTAAF 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Db         62 QYGNRAFTKYPRDIADYFKQSFPEGYSWERSMTYEDQGICIIKSDIRMEGDCFIYEIRYD 121

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180

Db        122 GVNFPPSGPVMQKKTLKWEPSTEKMYVRDGVLKGDVNMALLLEGGGHYRCDFRSTYKAKK 181

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIAR 215

Db        182 RVQLPDYHFVDHRIEILSHDNDYNTVKLSENAEAR 216


RESULT 8
Q7Z0W5
ID   Q7Z0W5      PRELIMINARY;      PRT;   225 AA.
AC   Q7Z0W5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cyan fluorescent protein.
OS   Montastraea cavernosa (great star coral).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC   Faviina; Faviidae; Montastraea.
OX   NCBI_TaxID=63558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=mc5;
RX   MEDLINE=22689801; PubMed=12777529;
RA   Kelmanson I.V., Matz M.V.;
RT   "Molecular Basis and Evolutionary Origins of Color Diversity in Great
RT   Star Coral Montastraea cavernosa (Scleractinia: Faviida).";
RL   Mol. Biol. Evol. 20:1125-1133(2003).
DR   EMBL; AY181556; AAO61602.1; -.
SQ   SEQUENCE   225 AA;  25843 MW;  13708587B7D93E35 CRC64;

  Query Match             52.4%;  Score 674;  DB 5;  Length 225;
  Best Local Similarity   54.9%;  Pred. No. 3.7e-55;
  Matches  118;  Conservative   42;  Mismatches   55;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60

Db          2 SVIKSVMKIKLHMDGIVNGHKFMITGEGEGKPFEGTHTIILKVKEGGPLPFAYDILTTAF 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Db         62 QYGNRVFTKYPKDIPDYFKQSFPEGYSWERSMTFEDQGVCTVTSDIKLEGDCFFYEIRFY 121

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180

Db        122 GVNFPSSGPVMQKKTLKWEPSTENMYVRDGVLLGDVSRTLLLEGDKHHRCNFRSTYRAKK 181

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIAR 215

Db        182 GVVLPEYHFVDHRIEILSHDKDYNTVEVYENAVAR 216


RESULT 9
Q9U6Y3
ID   Q9U6Y3      PRELIMINARY;      PRT;   266 AA.
AC   Q9U6Y3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Fluorescent protein FP484.
OS   Clavularia sp.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Alcyonaria; Alcyonacea;
OC   Clavulariidae; Clavularia.
OX   NCBI_TaxID=86521;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99436614; PubMed=10504696;
RA   Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
RA   Markelov M.L., Lukyanov S.A.;
RT   "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL   Nat. Biotechnol. 17:969-973(1999).
DR   EMBL; AF168424; AAF03374.1; -.
DR   GO; GO:0006091; P:energy pathways; IEA.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   266 AA;  30450 MW;  B4E97406E2708854 CRC64;

  Query Match             52.1%;  Score 671;  DB 5;  Length 266;
  Best Local Similarity   55.6%;  Pred. No. 8.6e-55;
  Matches  119;  Conservative   37;  Mismatches   58;  Indels    0;  Gaps    0;

Qy          2 VIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQCQ 61

Db         45 VIKPDMKIKLKMEGNVNGHAFVIEGEGEGKPYDGTHTLNLEVKEGAPLPFSYDILSNAFQ 104

Qy         62 YGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFSG 121

Db        105 YGNRALTKYPDDIADYFKQSFPEGYSWERTMTFEDKGIVKVKSDISMEEDSFIYEIRFDG 164

Qy        122 LNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKKP 181

Db        165 MNFPPNGPVMQKKTLKWEPSTEIMYVRDGVLVGDISHSLLLEGGGHYRCDFKSIYKAKKV 224

Qy        182 VKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIAR 215

Db        225 VKLPDYHFVDHRIEILNHDKDYNKVTLYENAVAR 258


RESULT 10
Q7Z0W6
ID   Q7Z0W6      PRELIMINARY;      PRT;   227 AA.
AC   Q7Z0W6;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Green fluorescent protein.
OS   Montastraea cavernosa (great star coral).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC   Faviina; Faviidae; Montastraea.
OX   NCBI_TaxID=63558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=mc4;
RX   MEDLINE=22689801; PubMed=12777529;
RA   Kelmanson I.V., Matz M.V.;
RT   "Molecular Basis and Evolutionary Origins of Color Diversity in Great
RT   Star Coral Montastraea cavernosa (Scleractinia: Faviida).";
RL   Mol. Biol. Evol. 20:1125-1133(2003).
DR   EMBL; AY181555; AAO61601.1; -.
SQ   SEQUENCE   227 AA;  26055 MW;  4BE2CB64FDB0E890 CRC64;

  Query Match             52.0%;  Score 669.5;  DB 5;  Length 227;
  Best Local Similarity   54.9%;  Pred. No. 9.8e-55;
  Matches  117;  Conservative   41;  Mismatches   52;  Indels    3;  Gaps    1;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60

Db          2 SVIKPDMKIKLRMEGAVNGHKFVIEGDGKGKPFEGKQTMDLTVIEGAPLPFAYDILTTVF 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQ---GNCFIYHV 117

Db         62 DYGNRVFAKYPKDIPDYFKQTFPEGYSWERSMTYEDQGICIATNDITMMKGVDDCFLYKI 121

Qy        118 KFSGLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYK 177
```

```
Db        122 RFDGVNFPANGPVMQRKTLKWEPSTEKMYVRDGVLKGDVNMALLLEGGGHYRCDFKTTYK 181

Qy        178 AKKPVKMPGYHYVDRKLDVTNHNKDYTSVEQCE 210
              ||| |::| ||:|| :::: :|:||| |:  |
Db        182 AKKVVQLPDYHFVDHRIEIVSHDKDYNKVKLYE 214


RESULT 11
Q7Z0W9
ID   Q7Z0W9      PRELIMINARY;      PRT;   225 AA.
AC   Q7Z0W9;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Red fluorescent protein.
OS   Montastraea cavernosa (great star coral).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC   Faviina; Faviidae; Montastraea.
OX   NCBI_TaxID=63558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=mc1;
RX   MEDLINE=22689801; PubMed=12777529;
RA   Kelmanson I.V., Matz M.V.;
RT   "Molecular Basis and Evolutionary Origins of Color Diversity in Great
RT   Star Coral Montastraea cavernosa (Scleractinia: Faviida).";
RL   Mol. Biol. Evol. 20:1125-1133(2003).
DR   EMBL; AY181552; AAO61598.1; -.
SQ   SEQUENCE   225 AA;  25822 MW;  830871F6936953E2 CRC64;

  Query Match             51.9%;  Score 667.5;  DB 5;  Length 225;
  Best Local Similarity   55.5%;  Pred. No. 1.5e-54;
  Matches  117;  Conservative   40;  Mismatches   53;  Indels    1;  Gaps    1;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||   |  |: | |:|||| | : |:|:|||||| |:: | | :| |||||:||::
Db          2 SVIKSVMKIKLRMEGSVNGHNFVIVGEGEGKPYEGTQSMDLTVKEGAPLPFAYDIMTTVF 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
               ||:  | |||: |||| || ||  |:||| |||||| :||  |: :::|:||   |:|
Db         62 HYGNRVFAKYPKHIPDYFKQMFPEGYSWERSMNFEDGGICTARNEITMEGDCFFNKVRFD 121

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              |:||||||||||||  |||:||::: |||:| |:  ||| ||||||| |:|::||:|||
Db        122 GVNFPPNGPVMQKKTLKWEPSTEKMYVRDGVLTGDINMALLLEGGGHYRCDFRTTYRAKK 181

Qy        181 P-VKMPGYHYVDRKLDVTNHNKDYTSVEQCE 210
               ||:| ||:||  :::  |:|:|| |:  |
Db        182 KGVKLPDYHFVDHSIEILRHDKEYTEVKLYE 212


RESULT 12
Q962P9
ID   Q962P9      PRELIMINARY;      PRT;   227 AA.
AC   Q962P9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Green fluorescent protein.
OS   Montastraea faveolata.
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC   Faviina; Faviidae; Montastraea.
OX   NCBI_TaxID=48498;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lesser M.P., Barry T.M., Mazel C., Matz M.V., Lukyanov S.A.,
RA   Falkowski P., Gorbunov M., Kolber Z.;
RT   "Green Fluorescent Proteins in Caribbean Scleractinian Corals.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF401282; AAK83923.1; -.
DR   GO; GO:0006091; P:energy pathways; IEA.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   227 AA;  26017 MW;  5E312C54EA47F589 CRC64;

  Query Match             51.9%;  Score 667.5;  DB 5;  Length 227;
  Best Local Similarity   54.9%;  Pred. No. 1.5e-54;
  Matches  117;  Conservative   40;  Mismatches   53;  Indels    3;  Gaps    1;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||   |  |: | | |||| | :|||||||:|| |:: | | :| |||||:|||:
Db          2 SVIKPDMKIKLRMEGAVNGHKFVIEGDGKGKPFEGTQSMDLTVKEGAPLPFAYDILTTVF 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQ---GNCFIYHV 117
               ||:  | |||:|||| ||:||  |:||| | :||  :|  :|| ::     :||:| :
Db         62 DYGNRVFAKYPQDIPDYFKQTFPEGYSWERSMTYEDQGICVATNDITLMKGVDDCFVYKI 121

Qy        118 KFSGLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYK 177
              :| |:||| |||||||||  |||:||::: |||:| |:  ||| ||||||| |:||:|||
Db        122 RFDGVNFPANGPVMQKKTLKWEPSTEKMYVRDGVLKGDVNMALLLEGGGHYRCDFKTTYK 181

Qy        178 AKKPVKMPGYHYVDRKLDVTNHNKDYTSVEQCE 210
              ||| |::| ||:|| :::: :|:|||  |:  |
Db        182 AKKFVQLPDYHFVDHRIEILSHDKDYNKVKLYE 214


RESULT 13
Q7Z0W8
ID   Q7Z0W8      PRELIMINARY;      PRT;   227 AA.
AC   Q7Z0W8;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Green fluorescent protein.
OS   Montastraea cavernosa (great star coral).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC   Faviina; Faviidae; Montastraea.
OX   NCBI_TaxID=63558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=mc2;
RX   MEDLINE=22689801; PubMed=12777529;
RA   Kelmanson I.V., Matz M.V.;
RT   "Molecular Basis and Evolutionary Origins of Color Diversity in Great
RT   Star Coral Montastraea cavernosa (Scleractinia: Faviida).";
RL   Mol. Biol. Evol. 20:1125-1133(2003).
DR   EMBL; AY181553; AAO61599.1; -.
SQ   SEQUENCE   227 AA;  26017 MW;  5E312C54EA47F589 CRC64;

  Query Match             51.9%;  Score 667.5;  DB 5;  Length 227;
  Best Local Similarity   54.9%;  Pred. No. 1.5e-54;
  Matches  117;  Conservative   40;  Mismatches   53;  Indels    3;  Gaps    1;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||   |  |: | | |||| | :|||||||:|| |:: | | :| |||||:|||:
Db          2 SVIKPDMKIKLRMEGAVNGHKFVIEGDGKGKPFEGTQSMDLTVKEGAPLPFAYDILTTVF 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQ---GNCFIYHV 117
               ||:  | |||:|||| ||:||  |:||| | :||  :|  :|| ::     :||:| :
Db         62 DYGNRVFAKYPQDIPDYFKQTFPEGYSWERSMTYEDQGICVATNDITLMKGVDDCFVYKI 121

Qy        118 KFSGLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYK 177
              :| |:||| |||||||||  |||:||::: |||:| |:  ||| ||||||| |:||:|||
Db        122 RFDGVNFPANGPVMQKKTLKWEPSTEKMYVRDGVLKGDVNMALLLEGGGHYRCDFKTTYK 181

Qy        178 AKKPVKMPGYHYVDRKLDVTNHNKDYTSVEQCE 210
              ||| |::| ||:|| :::: :|:|||  |:  |
Db        182 AKKFVQLPDYHFVDHRIEILSHDKDYNKVKLYE 214


RESULT 14
Q95UA7
ID   Q95UA7      PRELIMINARY;      PRT;   225 AA.
AC   Q95UA7;
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cyan fluorescent protein (Fragment).
OS   Montastraea cavernosa (great star coral).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC   Faviina; Faviidae; Montastraea.
OX   NCBI_TaxID=63558;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Falkowski P.G., Sun Y.;
RT   "Montastraea cavernosa fluorescent protein.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY056460; AAL17905.1; -.
DR   GO; GO:0006091; P:energy pathways; IEA.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
FT   NON_TER     225    225
SQ   SEQUENCE   225 AA;  25775 MW;  52DE2F716D083524 CRC64;

  Query Match             51.8%;  Score 666.5;  DB 5;  Length 225;
  Best Local Similarity   54.3%;  Pred. No. 1.9e-54;
  Matches  121;  Conservative   40;  Mismatches   59;  Indels    3;  Gaps    1;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||   |  |: | | |||| | : |:|:|||:||  |: | | :|||||||:|||:
Db          2 SVIKSVMKIKLRMDGIVNGHKFMITGEGEGKPFEGTHTIILKVKEGGPLPFAYDILTTAF 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||:  |||||:|||||  |||||   |:||| | |||  ||||::|  ::|:|| | ::|
Db         62 QYGNRVFTKYPKDIPDYFKQSFPEGYSWERSMTFEDQGVCTVTSDIKLEGDCFFYEIRFY 121

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              |:||| :||||||||  |||:|| :: |||:|:|:    | |||  |: | |:||| |||
Db        122 GVNFPSSGPVMQKKTLKWEPSTENMYVRDGVLLGDVSRTLLLEGDKHHRCNFRSTYGAKK 181

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARK---PVVA 220
                | :| ||:|| :::: :|:||| :||  | ::||    || |
Db        182 GVVLPEYHFVDHRIEILSHDKDYNTVEVYENAVARPSMLPVKA 224


RESULT 15
Q8T5F1
ID   Q8T5F1      PRELIMINARY;      PRT;   225 AA.
AC   Q8T5F1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   McavFP_7.5.
OS   Montastraea cavernosa (great star coral).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC   Faviina; Faviidae; Montastraea.
OX   NCBI_TaxID=63558;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21927629; PubMed=11929996;
RA   Labas Y.A., Gurskaya N.G., Yanushevich Y.G., Fradkov A.F.,
RA   Lukyanov K.A., Lukyanov S.A., Matz M.V.;
RT   "Diversity and evolution of the green fluorescent protein family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:4256-4261(2002).
DR   EMBL; AY037770; AAK71336.1; -.
DR   GO; GO:0006091; P:energy pathways; IEA.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR000786; Green_fl_protein.
DR   Pfam; PF01353; GFP; 1.
DR   PRINTS; PR01229; GFLUORESCENT.
DR   ProDom; PD013756; Green_fl_protein; 1.
SQ   SEQUENCE   225 AA;  25866 MW;  820C89437F8BDB32 CRC64;

  Query Match             51.0%;  Score 656.5;  DB 5;  Length 225;
  Best Local Similarity   54.5%;  Pred. No. 1.6e-53;
  Matches  115;  Conservative   40;  Mismatches   55;  Indels    1;  Gaps    1;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||   |  |: | |:|||| | : |:|:|||||| |:: | | :| |||||:||::
Db          2 SVIKSVMKIKLRMEGSVNGHNFVIVGEGEGKPYEGTQSMDLTVKEGAPLPFAYDIMTTVF 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
               ||:  | |||: |||| || ||  |:||| |||| | :||  |: :::|:||   |:|
Db         62 HYGNRVFAKYPKHIPDYFKQMFPEEYSWERSMNFEGGGICTARNEITMEGDCFFNKVRFD 121

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
              |:||||||||||||  |||:||::: |||:| |:  ||| ||||||| |:|::||:|||
Db        122 GVNFPPNGPVMQKKTLKWEPSTEKMYVRDGVLTGDINMALLLEGGGHYRCDFRTTYRAKK 181

Qy        181 P-VKMPGYHYVDRKLDVTNHNKDYTSVEQCE 210
               ||:| ||: |  :::  |:|:|| |:  |
Db        182 KGVKLPDYHFEDHSIEILRHDKEYTEVKLYE 212


Search completed: August 12, 2004, 06:19:37
Job time : 61.4078 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 87.5717 Seconds
(without alignments)
745.314 Million cell updates/sec

Title: US-09-890-463-3
Perfect score: 1268
Sequence: 1 SVIAKQMTYKVYMSGTVNGH.....SIARKPLVACCFFRVKSRHK 231

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1268	100.0	231	3	AAY97149	Aay97149 Pigment p
2	1268	100.0	231	5	ABP70025	Abp70025 Colour Fa
3	1242	97.9	235	3	AAY97150	Aay97150 Pigment p
4	1242	97.9	235	5	ABP70042	Abp70042 Colour Fa
5	1212	95.6	235	5	ABP70026	Abp70026 Colour Fa
6	1202	94.8	220	5	ABP70037	Abp70037 Colour Fa
7	1184	93.4	226	5	ABP70036	Abp70036 Colour Fa
8	1170	92.3	220	5	ABP69941	Abp69941 Colour Fa
9	1170	92.3	220	5	ABP69940	Abp69940 Colour Fa
10	1167	92.0	220	5	ABP69939	Abp69939 Colour Fa
11	1164	91.8	220	5	ABP69925	Abp69925 Colour Fa
12	1162	91.6	220	5	ABP69930	Abp69930 Colour Fa
13	1162	91.6	223	5	ABP70032	Abp70032 Colour Fa
14	1156	91.2	220	5	ABP69926	Abp69926 Colour Fa
15	1156	91.2	223	5	ABP70030	Abp70030 Colour Fa
16	1154	91.0	220	5	ABP69956	Abp69956 Colour Fa
17	1153	90.9	220	5	ABP69928	Abp69928 Colour Fa
18	1151	90.8	221	5	ABP69992	Abp69992 Colour Fa
19	1151	90.8	221	5	ABP69991	Abp69991 Colour Fa
20	1151	90.8	223	5	ABP70029	Abp70029 Colour Fa
21	1149.5	90.7	222	5	ABP70028	Abp70028 Colour Fa
22	1148	90.5	220	5	ABP70007	Abp70007 Colour Fa
23	1147	90.5	221	5	ABP69967	Abp69967 Colour Fa
24	1147	90.5	221	5	ABP69966	Abp69966 Colour Fa
25	1147	90.5	221	5	ABP70004	Abp70004 Colour Fa
26	1147	90.5	223	5	ABP70033	Abp70033 Colour Fa
27	1147	90.5	235	5	ABP69963	Abp69963 Colour Fa
28	1147	90.5	235	5	ABP69961	Abp69961 Colour Fa
29	1146	90.4	221	5	ABP69978	Abp69978 Colour Fa
30	1143	90.1	220	5	ABP69952	Abp69952 Colour Fa
31	1143	90.1	220	5	ABP69959	Abp69959 Colour Fa
32	1143	90.1	220	5	ABP69934	Abp69934 Colour Fa
33	1143	90.1	220	5	ABP69958	Abp69958 Colour Fa
34	1143	90.1	220	5	ABP69937	Abp69937 Colour Fa
35	1143	90.1	220	5	ABP69935	Abp69935 Colour Fa
36	1143	90.1	221	5	ABP69972	Abp69972 Colour Fa
37	1141.5	90.0	222	5	ABP70027	Abp70027 Colour Fa
38	1140	89.9	220	5	ABP69965	Abp69965 Colour Fa
39	1140	89.9	221	5	ABP70002	Abp70002 Colour Fa
40	1139	89.8	220	5	ABP69938	Abp69938 Colour Fa
41	1139	89.8	223	5	ABP70031	Abp70031 Colour Fa
42	1138	89.7	220	5	ABP69964	Abp69964 Colour Fa
43	1137	89.7	220	5	ABP69993	Abp69993 Colour Fa
44	1137	89.7	220	5	ABP69989	Abp69989 Colour Fa
45	1137	89.7	221	5	ABP69986	Abp69986 Colour Fa

ALIGNMENTS

RESULT 1
AAY97149
ID AAY97149 standard; protein; 231 AA.
XX
AC AAY97149;
XX
DT 04-DEC-2000 (first entry)
XX
DE Pigment protein from coral tissue POC3.
XX
KW N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW UV filter; POC3.
XX
OS Acropora aspera.
XX
FH Key Location/Qualifiers
FT Misc-difference 61. .63
FT /label= Chromophore_motif
FT Misc-difference 158
FT /note= "critical residue in the vicinity of the
FT fluorophore"
FT Misc-difference 192
FT /note= "critical residue in the vicinity of the
FT fluorophore"
FT Misc-difference 210
FT /note= "critical residue in the vicinity of the
FT fluorophore"
XX
PN WO200046233-A1.
XX
PD 10-AUG-2000.
XX
PF 02-FEB-2000; 2000WO-AU000056.
XX
PR 02-FEB-1999; 99AU-00008463.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Hoegh-Guldberg O, Dove S;
XX
DR WPI; 2000-532892/48.
DR N-PSDB; AAA52082.
XX
PT Novel pigment protein derived from corals capable of emitting
PT fluorescence upon irradiation by incident light useful as tissue marker,
PT fluorescent marker or general dyestuff.
XX

```
PS   Claim 13; Page 42-43; 49pp; English.
XX
CC   cDNA libraries were constructed from a blue pigmented coral, Acropora
CC   aspera to isolate sequences encoding polypeptides with N-terminal
CC   sequences as shown in AAY97147-48. Pigment protein from coral tissue
CC   (PPCT) is capable of emitting fluorescence upon irradiation by incident
CC   light whose maximal absorbance is in the range of 320-600 nm and a
CC   maximal fluorescence emission is in the range of 300-700 nm. PPCT may be
CC   used as a tissue marker, fluorescent marker (e.g. to follow gene
CC   expression in transformed tissues) or general dyestuff (all claimed).
CC   PPCT may also be used in sunscreen formulations or UV filters (both
CC   claimed)
XX
SQ   Sequence 231 AA;

  Query Match             100.0%;  Score 1268;  DB 3;  Length 231;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-128;
  Matches  231;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVACCFFRVKSRHK 231
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVACCFFRVKSRHK 231


RESULT 2
ABP70025
ID   ABP70025 standard; protein; 231 AA.
XX
AC   ABP70025;
XX
DT   06-AUG-2003  (revised)
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 201.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 6; Page 478; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 231 AA;

  Query Match             100.0%;  Score 1268;  DB 5;  Length 231;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-128;
  Matches  231;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVACCFFRVKSRHK 231
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVACCFFRVKSRHK 231


RESULT 3
AAY97150
ID   AAY97150 standard; protein; 235 AA.
XX
AC   AAY97150;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Pigment protein from coral tissue POC4.
XX
KW   N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW   tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW   UV filter; POC3.
XX
OS   Acropora aspera.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 61. .63
FT                   /label= Chromophore_motif
FT   Misc-difference 158
FT                   /note= "critical residue in the vicinity of the
```

```
FT                   fluorophore"
FT   Misc-difference 192
FT                   /note= "critical residue in the vicinity of the
FT                   fluorophore"
FT   Misc-difference 210
FT                   /note= "critical residue in the vicinity of the
FT                   fluorophore"
XX
PN   WO200046233-A1.
XX
PD   10-AUG-2000.
XX
PF   02-FEB-2000; 2000WO-AU000056.
XX
PR   02-FEB-1999;   99AU-00008463.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Hoegh-Guldberg O,  Dove S;
XX
DR   WPI; 2000-532892/48.
DR   N-PSDB; AAA52083.
XX
PT   Novel pigment protein derived from corals capable of emitting
PT   fluorescence upon irradiation by incident light useful as tissue marker,
PT   fluorescent marker or general dyestuff.
XX
PS   Claim 13; Page 43-44; 49pp; English.
XX
CC   cDNA libraries were constructed from a blue pigmented coral, Acropora
CC   aspera to isolate sequences encoding polypeptides with N-terminal
CC   sequences as shown in AAY97147-48. Pigment protein from coral tissue
CC   (PPCT) is capable of emitting fluorescence upon irradiation by incident
CC   light whose maximal absorbance is in the range of 320-600 nm and a
CC   maximal fluorescence emission is in the range of 300-700 nm. PPCT may be
CC   used as a tissue marker, fluorescent marker (e.g. to follow gene
CC   expression in transformed tissues) or general dyestuff (all claimed).
CC   PPCT may also be used in sunscreen formulations or UV filters (both
CC   claimed)
XX
SQ   Sequence 235 AA;

  Query Match             97.9%;  Score 1242;  DB 3;  Length 235;
  Best Local Similarity   98.3%;  Pred. No. 8.9e-126;
  Matches  227;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVACCFFRVKSRHK 231
              |||||||||||||||||||||||||| |||||||:||| |||||||||
Db        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVACRFFRVKSRHK 231


RESULT 4
ABP70042
ID   ABP70042 standard; protein; 235 AA.
XX
AC   ABP70042;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 245.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Acropora aspera.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Example 20; Page 502-503; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences
XX
SQ   Sequence 235 AA;

  Query Match             97.9%;  Score 1242;  DB 5;  Length 235;
  Best Local Similarity   98.3%;  Pred. No. 8.9e-126;
  Matches  227;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180
```

```
Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVACCFFRVKSRHK 231
              ||||||||||||||||||||||||||||| |||||||||:||| |||||||||
Db        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVACRFFRVKSRHK 231
```

RESULT 5
ABP70026
ID ABP70026 standard; protein; 235 AA.
XX
AC ABP70026;
XX
DT 06-AUG-2003 (revised)
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 202.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Unidentified.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 479; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 235 AA;

Query Match 95.6%; Score 1212; DB 5; Length 235;
Best Local Similarity 96.5%; Pred. No. 1.6e-122;
Matches 223; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

```
Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||||||||||||||||||||||||||||| |||| ||||||||||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGLPYEGGQTVRLAVTKGGPLPFAWDILSPQC 60

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKRS 120

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVACCFFRVKSRHK 231
              |||||||||||||||||| ||||||| |||||||:||| |||||||||
Db        181 PVKMPGYHYVDRKLDVTNHNLDYTSVEQCEISIARKPVVACRFFRVKSRHK 231
```

RESULT 6
ABP70037
ID ABP70037 standard; protein; 220 AA.
XX
AC ABP70037;
XX
DT 06-AUG-2003 (revised)
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 239.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Unidentified.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Example 19; Page 497-498; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other

CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 220 AA;

Query Match 94.8%; Score 1202; DB 5; Length 220;
Best Local Similarity 100.0%; Pred. No. 1.7e-121;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
Db 1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60

Qy 61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
Db 61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy 121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180
Db 121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180

Qy 181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVA 220
Db 181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVA 220

RESULT 7
ABP70036
ID ABP70036 standard; protein; 226 AA.
XX
AC ABP70036;
XX
DT 06-AUG-2003 (revised)
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 238.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Unidentified.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Example 19; Page 496-497; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 226 AA;

Query Match 93.4%; Score 1184; DB 5; Length 226;
Best Local Similarity 96.0%; Pred. No. 1.6e-119;
Matches 217; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
Db 1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLTVTKGGPLPFAWDILSPQS 60

Qy 61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
Db 61 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy 121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180
Db 121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180

Qy 181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVACCFFRV 226
Db 181 PVRMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVAWCFFRV 226

RESULT 8
ABP69941
ID ABP69941 standard; protein; 220 AA.
XX
AC ABP69941;
XX
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 54.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Millepora sp.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.

XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 330-331; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 220 AA;

Query Match 92.3%; Score 1170; DB 5; Length 220;
Best Local Similarity 97.7%; Pred. No. 5.1e-118;
Matches 215; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLTVTKGGPLPFAWDILSPQS 60

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db         61 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVA 220
              |||||||||||||||||||||||||||||||||||||:||
Db        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPVVA 220
```

RESULT 9
ABP69940
ID ABP69940 standard; protein; 220 AA.
XX
AC ABP69940;

XX
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 52.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Millepora sp.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 327-328; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 220 AA;

Query Match 92.3%; Score 1170; DB 5; Length 220;
Best Local Similarity 97.7%; Pred. No. 5.1e-118;
Matches 215; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLTVTKGGPLPFAWDILSPQS 60

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db         61 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
```

```
Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVA 220
              |||||||||||||||||||||||||||||||||||||:||
Db        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPVVA 220


RESULT 10
ABP69939
ID   ABP69939 standard; protein; 220 AA.
XX
AC   ABP69939;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 50.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Millepora sp.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 325-326; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences
XX
SQ   Sequence 220 AA;

  Query Match             92.0%;  Score 1167;  DB 5;  Length 220;
  Best Local Similarity   97.3%;  Pred. No. 1.1e-117;
  Matches  214;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||||||||||||||||||||||||||||||||||||:| ||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVKLTVTKGGPLPFAWDILSPQS 60

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db         61 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVA 220
              |||||||||||||||||||||||||||||||||||||:||
Db        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPVVA 220


RESULT 11
ABP69925
ID   ABP69925 standard; protein; 220 AA.
XX
AC   ABP69925;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 22.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Acropora aspera.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 286-287; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
```

CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 220 AA;

Query Match 91.8%; Score 1164; DB 5; Length 220;
Best Local Similarity 97.3%; Pred. No. 2.3e-117;
Matches 214; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLTVTKGGPLPFAWDILSPQS 60

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db         61 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLLARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVA 220
              |||||||||||||||||||||||||||||||||||||:||
Db        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPVVA 220
```

RESULT 12
ABP69930
ID ABP69930 standard; protein; 220 AA.
XX
AC ABP69930;
XX
DT 06-AUG-2003 (revised)
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 32.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Caulastrea sp.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 298-299; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 220 AA;

Query Match 91.6%; Score 1162; DB 5; Length 220;
Best Local Similarity 97.3%; Pred. No. 3.8e-117;
Matches 214; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLTVTKGGPLPFAWDILSPQS 60

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db         61 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLSARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVA 220
              |||||||||||||||||||||||||||||||||||||:||
Db        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPVVA 220
```

RESULT 13
ABP70032
ID ABP70032 standard; protein; 223 AA.
XX
AC ABP70032;
XX
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 220.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Sinularia sp.
XX

```
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Disclosure; Page 489; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences
XX
SQ   Sequence 223 AA;

  Query Match             91.6%;  Score 1162;  DB 5;  Length 223;
  Best Local Similarity   97.3%;  Pred. No. 3.8e-117;
  Matches  214;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              |||||||||| |||||||||||  :||||||||||||||||||||||||||||||||||||
Db         62 QYGSIPFTKYLEDIPDYVKQSFPEGFTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 121

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        122 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 181

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVA 220
              |||||||||||||||||||||||||| ||||||||||
Db        182 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPLVA 221


RESULT 14
ABP69926
ID   ABP69926 standard; protein; 220 AA.
XX
AC   ABP69926;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 24.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Acropora aspera.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 289; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences
XX
SQ   Sequence 220 AA;

  Query Match             91.2%;  Score 1156;  DB 5;  Length 220;
  Best Local Similarity   96.4%;  Pred. No. 1.7e-116;
  Matches  212;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLTVTKGGPLPFAWDILSPQS 60
```

```
Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              ||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db         61 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 180

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVA 220
              ||:|||||||||||||||||||||||| ||||||||:||
Db        181 PVRMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 220


RESULT 15
ABP70030
ID   ABP70030 standard; protein; 223 AA.
XX
AC   ABP70030;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 216.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Tubastrea sp.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Example 18; Page 486; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences
XX
SQ   Sequence 223 AA;

  Query Match             91.2%;  Score 1156;  DB 5;  Length 223;
  Best Local Similarity   96.4%;  Pred. No. 1.7e-116;
  Matches  212;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLAVTKGGPLPFAWDILSPQC 60
              ||||||||||||||||||||||||||||||||||||||||||:| ||||||||||||||||
Db          2 SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVKLTVTKGGPLPFAWDILSPQS 61

Qy         61 QYGSIPFTKYPEDIPDYVKQSFPGRYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 120
              ||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db         62 QYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTVSNDSSIQGNCFIYHVKFS 121

Qy        121 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKARK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        122 GLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMALKLEGGGHYLCEFKSTYKAKK 181

Qy        181 PVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPLVA 220
              ||||||||||||||||||||||||| ||||||||:||
Db        182 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPVVA 221


Search completed: August 12, 2004, 06:17:05
Job time : 88.5717 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 6.44467 Seconds
(without alignments)
745.314 Million cell updates/sec

Title: US-09-890-463-2
Perfect score: 83
Sequence: 1 SVIAKQMTYKVYMSGTV 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	83	100.0	17	3	AAY97148	Aay97148 Pigment p
2	83	100.0	18	3	AAY97151	Aay97151 Pigment p
3	83	100.0	25	3	AAY97152	Aay97152 Pigment p
4	83	100.0	169	5	ABP69949	Abp69949 Colour Fa
5	83	100.0	169	5	ABP69944	Abp69944 Colour Fa
6	83	100.0	200	5	ABP69957	Abp69957 Colour Fa
7	83	100.0	220	5	ABP69941	Abp69941 Colour Fa
8	83	100.0	220	5	ABP69952	Abp69952 Colour Fa
9	83	100.0	220	5	ABP69925	Abp69925 Colour Fa
10	83	100.0	220	5	ABP69947	Abp69947 Colour Fa
11	83	100.0	220	5	ABP69959	Abp69959 Colour Fa
12	83	100.0	220	5	ABP69940	Abp69940 Colour Fa
13	83	100.0	220	5	ABP69943	Abp69943 Colour Fa
14	83	100.0	220	5	ABP69955	Abp69955 Colour Fa
15	83	100.0	220	5	ABP69929	Abp69929 Colour Fa
16	83	100.0	220	5	ABP69934	Abp69934 Colour Fa
17	83	100.0	220	5	ABP69958	Abp69958 Colour Fa
18	83	100.0	220	5	ABP69939	Abp69939 Colour Fa
19	83	100.0	220	5	ABP69953	Abp69953 Colour Fa
20	83	100.0	220	5	ABP69938	Abp69938 Colour Fa
21	83	100.0	220	5	ABP69945	Abp69945 Colour Fa
22	83	100.0	220	5	ABP69927	Abp69927 Colour Fa
23	83	100.0	220	5	ABP69946	Abp69946 Colour Fa
24	83	100.0	220	5	ABP69926	Abp69926 Colour Fa
25	83	100.0	220	5	ABP69956	Abp69956 Colour Fa
26	83	100.0	220	5	ABP69937	Abp69937 Colour Fa
27	83	100.0	220	5	ABP69932	Abp69932 Colour Fa
28	83	100.0	220	5	ABP69928	Abp69928 Colour Fa
29	83	100.0	220	5	ABP69931	Abp69931 Colour Fa
30	83	100.0	220	5	ABP69935	Abp69935 Colour Fa
31	83	100.0	220	5	ABP69936	Abp69936 Colour Fa
32	83	100.0	220	5	ABP69948	Abp69948 Colour Fa
33	83	100.0	220	5	ABP70037	Abp70037 Colour Fa
34	83	100.0	220	5	ABP69930	Abp69930 Colour Fa
35	83	100.0	222	5	ABP70028	Abp70028 Colour Fa
36	83	100.0	222	5	ABP70027	Abp70027 Colour Fa
37	83	100.0	223	5	ABP70033	Abp70033 Colour Fa
38	83	100.0	223	5	ABP70030	Abp70030 Colour Fa
39	83	100.0	223	5	ABP70029	Abp70029 Colour Fa
40	83	100.0	223	5	ABP70031	Abp70031 Colour Fa
41	83	100.0	223	5	ABP70032	Abp70032 Colour Fa
42	83	100.0	226	5	ABP70036	Abp70036 Colour Fa
43	83	100.0	231	3	AAY97149	Aay97149 Pigment p
44	83	100.0	231	5	ABP70025	Abp70025 Colour Fa
45	83	100.0	235	3	AAY97150	Aay97150 Pigment p

ALIGNMENTS

RESULT 1
AAY97148
ID AAY97148 standard; peptide; 17 AA.
XX
AC AAY97148;
XX
DT 04-DEC-2000 (first entry)
XX
DE Pigment protein from coral tissue N-terminal peptide 2.
XX
KW N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW UV filter.
XX
OS Acropora horrida.
XX
PN WO200046233-A1.
XX
PD 10-AUG-2000.
XX
PF 02-FEB-2000; 2000WO-AU000056.
XX
PR 02-FEB-1999; 99AU-00008463.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Hoegh-Guldberg O, Dove S;
XX
DR WPI; 2000-532892/48.
XX
PT Novel pigment protein derived from corals capable of emitting
PT fluorescence upon irradiation by incident light useful as tissue marker,
PT fluorescent marker or general dyestuff.
XX
PS Claim 4; Page 42; 49pp; English.
XX
CC The N-terminal peptides shown in AAY97147-48 are from pigment protein
CC from coral tissue (PPCT). PPCT is capable of emitting fluorescence upon
CC irradiation by incident light whose maximal absorbance is in the range of
CC 320-600 nm and a maximal fluorescence emission is in the range of 300-700
CC nm. PPCT may be used as a tissue marker, fluorescent marker (e.g. to
CC follow gene expression in transformed tissues) or general dyestuff (all
CC claimed). PPCT may also be used in sunscreen formulations or UV filters
CC (both claimed)
XX
SQ Sequence 17 AA;

Query Match 100.0%; Score 83; DB 3; Length 17;

```
  Best Local Similarity   100.0%;  Pred. No. 3.6e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTV 17


RESULT 2
AAY97151
ID   AAY97151 standard; peptide; 18 AA.
XX
AC   AAY97151;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Pigment protein from coral tissue N-terminal peptide 3.
XX
KW   N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW   tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW   UV filter.
XX
OS   Acropora aspera.
OS   Montipora caliculata.
OS   Porites murrayensis.
XX
PN   WO200046233-A1.
XX
PD   10-AUG-2000.
XX
PF   02-FEB-2000; 2000WO-AU000056.
XX
PR   02-FEB-1999;   99AU-00008463.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Hoegh-Guldberg O,  Dove S;
XX
DR   WPI; 2000-532892/48.
XX
PT   Novel pigment protein derived from corals capable of emitting
PT   fluorescence upon irradiation by incident light useful as tissue marker,
PT   fluorescent marker or general dyestuff.
XX
PS   Example 2; Page 18; 49pp; English.
XX
CC   The N-terminal peptides shown in AAY97151-52 are from pigment protein
CC   from coral tissue (PPCT). PPCT is capable of emitting fluorescence upon
CC   irradiation by incident light whose maximal absorbance is in the range of
CC   320-600 nm and a maximal fluorescence emission is in the range of 300-700
CC   nm. PPCT may be used as a tissue marker, fluorescent marker (e.g. to
CC   follow gene expression in transformed tissues) or general dyestuff (all
CC   claimed). PPCT may also be used in sunscreen formulations or UV filters
CC   (both claimed)
XX
SQ   Sequence 18 AA;

  Query Match             100.0%;  Score 83;  DB 3;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTV 17


RESULT 3
AAY97152
ID   AAY97152 standard; peptide; 25 AA.
XX
AC   AAY97152;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Pigment protein from coral tissue N-terminal peptide 4.
XX
KW   N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW   tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW   UV filter.
XX
OS   Porites lobata.
XX
PN   WO200046233-A1.
XX
PD   10-AUG-2000.
XX
PF   02-FEB-2000; 2000WO-AU000056.
XX
PR   02-FEB-1999;   99AU-00008463.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Hoegh-Guldberg O,  Dove S;
XX
DR   WPI; 2000-532892/48.
XX
PT   Novel pigment protein derived from corals capable of emitting
PT   fluorescence upon irradiation by incident light useful as tissue marker,
PT   fluorescent marker or general dyestuff.
XX
PS   Example 2; Page 18; 49pp; English.
XX
CC   The N-terminal peptides shown in AAY97151-52 are from pigment protein
CC   from coral tissue (PPCT). PPCT is capable of emitting fluorescence upon
CC   irradiation by incident light whose maximal absorbance is in the range of
CC   320-600 nm and a maximal fluorescence emission is in the range of 300-700
CC   nm. PPCT may be used as a tissue marker, fluorescent marker (e.g. to
CC   follow gene expression in transformed tissues) or general dyestuff (all
CC   claimed). PPCT may also be used in sunscreen formulations or UV filters
CC   (both claimed)
XX
SQ   Sequence 25 AA;

  Query Match             100.0%;  Score 83;  DB 3;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTV 17


RESULT 4
ABP69949
ID   ABP69949 standard; protein; 169 AA.
XX
AC   ABP69949;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 70.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Platygyra sp.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
```

PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 349; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 169 AA;

Query Match 100.0%; Score 83; DB 5; Length 169;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
||||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

RESULT 5
ABP69944
ID ABP69944 standard; protein; 169 AA.
XX
AC ABP69944;
XX
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 60.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Porites murrayensis.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 337; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 169 AA;

Query Match 100.0%; Score 83; DB 5; Length 169;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
||||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

RESULT 6
ABP69957
ID ABP69957 standard; protein; 200 AA.
XX
AC ABP69957;
XX
DT 06-AUG-2003 (revised)
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 84.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Montipora sp.
XX

PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 363-364; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 200 AA;

Query Match 100.0%; Score 83; DB 5; Length 200;
Best Local Similarity 100.0%; Pred. No. 6.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

RESULT 7
ABP69941
ID ABP69941 standard; protein; 220 AA.
XX
AC ABP69941;
XX
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 54.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Millepora sp.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 330-331; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 220 AA;

Query Match 100.0%; Score 83; DB 5; Length 220;
Best Local Similarity 100.0%; Pred. No. 7e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

RESULT 8
ABP69952
ID ABP69952 standard; protein; 220 AA.
XX
AC ABP69952;
XX
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 74.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;

KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Platygyra sp.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 351-352; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 220 AA;

Query Match 100.0%; Score 83; DB 5; Length 220;
Best Local Similarity 100.0%; Pred. No. 7e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

RESULT 9
ABP69925
ID ABP69925 standard; protein; 220 AA.
XX
AC ABP69925;
XX
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 22.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Acropora aspera.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 286-287; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences
XX
SQ Sequence 220 AA;

Query Match 100.0%; Score 83; DB 5; Length 220;
Best Local Similarity 100.0%; Pred. No. 7e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

RESULT 10
ABP69947
ID ABP69947 standard; protein; 220 AA.
XX
AC ABP69947;

```
XX
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 66.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Platygyra sp.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 344-345; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences
XX
SQ   Sequence 220 AA;

  Query Match             100.0%;  Score 83;  DB 5;  Length 220;
  Best Local Similarity   100.0%;  Pred. No. 7e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTV 17


RESULT 11
ABP69959
```

```
ID   ABP69959 standard; protein; 220 AA.
XX
AC   ABP69959;
XX
DT   06-AUG-2003  (revised)
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 88.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Montipora sp.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 368-369; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 220 AA;

  Query Match             100.0%;  Score 83;  DB 5;  Length 220;
  Best Local Similarity   100.0%;  Pred. No. 7e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTV 17
```

```
RESULT 12
ABP69940
ID   ABP69940 standard; protein; 220 AA.
XX
AC   ABP69940;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 52.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Millepora sp.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 327-328; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences
XX
SQ   Sequence 220 AA;

  Query Match             100.0%;  Score 83;  DB 5;  Length 220;
  Best Local Similarity   100.0%;  Pred. No. 7e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTV 17


RESULT 13
ABP69943
ID   ABP69943 standard; protein; 220 AA.
XX
AC   ABP69943;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 58.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Porites murrayensis.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 335; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences
XX
SQ   Sequence 220 AA;

  Query Match             100.0%;  Score 83;  DB 5;  Length 220;
```

```
  Best Local Similarity   100.0%;  Pred. No. 7e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTV 17


RESULT 14
ABP69955
ID   ABP69955 standard; protein; 220 AA.
XX
AC   ABP69955;
XX
DT   06-AUG-2003  (revised)
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 80.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Pavona decussata.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 359; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   sequences given in records ABP69924-ABP70048 represent CFM related amino
CC   acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 220 AA;

  Query Match           100.0%;  Score 83;  DB 5;  Length 220;
  Best Local Similarity   100.0%;  Pred. No. 7e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAKQMTYKVYMSGTV 17
              |||||||||||||||||
Db          1 SVIAKQMTYKVYMSGTV 17


RESULT 15
ABP69929
ID   ABP69929 standard; protein; 220 AA.
XX
AC   ABP69929;
XX
DT   06-AUG-2003  (revised)
DT   22-JAN-2003  (first entry)
XX
DE   Colour Facilitating molecule (CFM) related sequence #SEQ ID 30.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Acanthastrea sp.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 5; Page 296-297; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
```

CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 220 AA;

Query Match 100.0%; Score 83; DB 5; Length 220;
Best Local Similarity 100.0%; Pred. No. 7e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAKQMTYKVYMSGTV 17
||||||||||||||||||
Db 1 SVIAKQMTYKVYMSGTV 17

Search completed: August 12, 2004, 06:17:04
Job time : 6.44467 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 1.89549 Seconds
(without alignments)
745.314 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	100.0	5	3	AAY97147	Aay97147 Pigment p
2	21	100.0	5	5	ABB99061	Abb99061 N-termina
3	21	100.0	10	4	AAG87969	Aag87969 Saccharom
4	21	100.0	10	4	AAG87968	Aag87968 Saccharom
5	21	100.0	13	5	ABP70008	Abp70008 Colour Fa
6	21	100.0	16	5	ABB99073	Abb99073 N-termina
7	21	100.0	16	5	ABB99074	Abb99074 N-termina
8	21	100.0	16	5	ABB99072	Abb99072 N-termina
9	21	100.0	17	3	AAY97148	Aay97148 Pigment p
10	21	100.0	18	3	AAY97151	Aay97151 Pigment p
11	21	100.0	25	3	AAY97152	Aay97152 Pigment p
12	21	100.0	50	4	AAB62002	Aab62002 C-myb pro
13	21	100.0	73	4	ABG14062	Abg14062 Novel hum
14	21	100.0	73	5	ABP34419	Abp34419 Human iso
15	21	100.0	75	6	ABP77483	Abp77483 N. gonorr
16	21	100.0	76	4	ABB12172	Abb12172 Human MEC
17	21	100.0	76	4	ABG16073	Abg16073 Novel hum
18	21	100.0	78	4	ABG16074	Abg16074 Novel hum
19	21	100.0	81	5	ABP39391	Abp39391 Staphyloc
20	21	100.0	89	3	AAG01464	Aag01464 Human sec
21	21	100.0	91	4	ABG16071	Abg16071 Novel hum
22	21	100.0	102	2	AAY29549	Aay29549 Human lun
23	21	100.0	102	3	AAB44472	Aab44472 Human lun
24	21	100.0	102	4	AAE13814	Aae13814 Human lun
25	21	100.0	102	7	ADD66504	Add66504 Human lun
26	21	100.0	102	7	ADE87758	Ade87758 Human lun
27	21	100.0	111	7	ADC95761	Adc95761 E. faeciu
28	21	100.0	137	6	ABM70003	Abm70003 Photorhab
29	21	100.0	144	4	ABG16072	Abg16072 Novel hum
30	21	100.0	144	4	ABG14060	Abg14060 Novel hum
31	21	100.0	147	5	ABP08878	Abp08878 Human ORF
32	21	100.0	158	6	ABM69564	Abm69564 Photorhab
33	21	100.0	169	5	ABP69949	Abp69949 Colour Fa
34	21	100.0	169	5	ABP69944	Abp69944 Colour Fa
35	21	100.0	180	4	ABG03673	Abg03673 Novel hum
36	21	100.0	182	4	ABG07747	Abg07747 Novel hum
37	21	100.0	188	6	ABM73297	Abm73297 Staphyloc
38	21	100.0	196	5	ABP65304	Abp65304 Bifidobac
39	21	100.0	200	5	ABP69957	Abp69957 Colour Fa
40	21	100.0	200	5	ABB47893	Abb47893 Listeria
41	21	100.0	220	5	ABP69941	Abp69941 Colour Fa
42	21	100.0	220	5	ABP69952	Abp69952 Colour Fa
43	21	100.0	220	5	ABP69925	Abp69925 Colour Fa
44	21	100.0	220	5	ABP69954	Abp69954 Colour Fa
45	21	100.0	220	5	ABP69947	Abp69947 Colour Fa

ALIGNMENTS

RESULT 1
AAY97147
ID AAY97147 standard; peptide; 5 AA.
XX
AC AAY97147;
XX
DT 04-DEC-2000 (first entry)
XX
DE Pigment protein from coral tissue N-terminal peptide 1.
XX
KW N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW UV filter.
XX
OS Acropora aspera.
OS Acropora horrida.
OS Montipora caliculata.
OS Montipora monasteriata.
OS Porites murrayensis.
OS Porites lobata.
XX
PN WO200046233-A1.
XX
PD 10-AUG-2000.
XX
PF 02-FEB-2000; 2000WO-AU000056.
XX
PR 02-FEB-1999; 99AU-00008463.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Hoegh-Guldberg O, Dove S;
XX
DR WPI; 2000-532892/48.
XX
PT Novel pigment protein derived from corals capable of emitting
PT fluorescence upon irradiation by incident light useful as tissue marker,
PT fluorescent marker or general dyestuff.
XX
PS Claim 3; Page 42; 49pp; English.
XX
CC The N-terminal peptides shown in AAY97147-48 are from pigment protein
CC from coral tissue (PPCT). PPCT is capable of emitting fluorescence upon
CC irradiation by incident light whose maximal absorbance is in the range of
CC 320-600 nm and a maximal fluorescence emission is in the range of 300-700
CC nm. PPCT may be used as a tissue marker, fluorescent marker (e.g. to
CC follow gene expression in transformed tissues) or general dyestuff (all
CC claimed). PPCT may also be used in sunscreen formulations or UV filters

```
CC   (both claimed)
XX
SQ   Sequence 5 AA;

  Query Match             100.0%;  Score 21;  DB 3;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db          1 SVIAK 5


RESULT 2
ABB99061
ID   ABB99061 standard; peptide; 5 AA.
XX
AC   ABB99061;
XX
DT   22-JAN-2003  (first entry)
XX
DE   N-terminal amino acid sequence of a CFM #1.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 3; Page 278; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 5 AA;

  Query Match             100.0%;  Score 21;  DB 5;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db          1 SVIAK 5


RESULT 3
AAG87969
ID   AAG87969 standard; peptide; 10 AA.
XX
AC   AAG87969;
XX
DT   11-SEP-2001  (first entry)
XX
DE   Saccharomyces cerevisiae peptide, SEQ ID NO: 2918.
XX
KW   Saccharomyces cerevisiae; complementary peptide; peptide identification;
KW   drug discovery; drug design.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO200142276-A1.
XX
PD   14-JUN-2001.
XX
PF   13-DEC-2000; 2000WO-GB004773.
XX
PR   13-DEC-1999;   99GB-00029471.
XX
PA   (PROT-) PROTEOM LTD.
XX
PI   Roberts GW,  Heal JR;
XX
DR   WPI; 2001-367863/38.
XX
PT   Identifying complementary peptides by analysis of protein and nucleotide
PT   sequence databases, useful in drug design.
XX
PS   Example 5; Page 432; 488pp; English.
XX
CC   The invention relates to the identification of complementary peptides by
CC   analysis of protein and nucleotide sequence databases from higher
CC   eukaryotic genomes, excluding human and plants. The specific
CC   complementary peptides interact with their relevant target proteins
CC   encoded in the eukaryote genome. The peptides may be used as reagents and
CC   drugs for drug discovery and as lead ligands for drug design and
CC   development. The present sequence is a complementary peptide from
CC   Saccharomyces cerevisiae
XX
SQ   Sequence 10 AA;

  Query Match             100.0%;  Score 21;  DB 4;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db          6 SVIAK 10


RESULT 4
AAG87968
ID   AAG87968 standard; peptide; 10 AA.
```

XX
AC AAG87968;
XX
DT 11-SEP-2001 (first entry)
XX
DE Saccharomyces cerevisiae peptide, SEQ ID NO: 2917.
XX
KW Saccharomyces cerevisiae; complementary peptide; peptide identification;
KW drug discovery; drug design.
XX
OS Saccharomyces cerevisiae.
XX
PN WO200142276-A1.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB004773.
XX
PR 13-DEC-1999; 99GB-00029471.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
DR WPI; 2001-367863/38.
XX
PT Identifying complementary peptides by analysis of protein and nucleotide
PT sequence databases, useful in drug design.
XX
PS Example 5; Page 432; 488pp; English.
XX
CC The invention relates to the identification of complementary peptides by
CC analysis of protein and nucleotide sequence databases from higher
CC eukaryotic genomes, excluding human and plants. The specific
CC complementary peptides interact with their relevant target proteins
CC encoded in the eukaryote genome. The peptides may be used as reagents and
CC drugs for drug discovery and as lead ligands for drug design and
CC development. The present sequence is a complementary peptide from
CC Saccharomyces cerevisiae
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 21; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
|||||
Db 6 SVIAK 10

RESULT 5
ABP70008
ID ABP70008 standard; peptide; 13 AA.
XX
AC ABP70008;
XX
DT 06-AUG-2003 (revised)
DT 22-JAN-2003 (first entry)
XX
DE Colour Facilitating molecule (CFM) related sequence #SEQ ID 184.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Pavona decussata.
XX
PN WO200270703-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-GB000928.
XX
PR 02-MAR-2001; 2001US-0273227P.
PR 21-MAR-2001; 2001AU-00003874.
PR 15-OCT-2001; 2001US-0329816P.
XX
PA (NUFA-) NUFARM LTD.
PA (UYQU) UNIV QUEENSLAND.
PA (JONE/) JONES E L.
XX
PI Jones EL, Karan M, Brugliera F, Mason J, Dove SG;
PI Hoegh-Guldberg IO, Prescott M;
XX
DR WPI; 2002-740765/80.
XX
PT Novel color-facilitating molecule for producing a biomatrix, has a
PT polypeptide which alone/along with molecules imparts altered visual
PT characteristics to cells in the absence of excitation by extraneous non-
PT white light.
XX
PS Claim 5; Page 473; 510pp; English.
XX
CC The invention relates to an isolated colour-facilitating molecule (CFM)
CC comprising a polypeptide which, in a cell, alone or together with one or
CC more other molecules imparts an altered visual characteristic to the cell
CC when visualised by a human eye in the absence of excitation by extraneous
CC non-white light or particle emission. CFMs are useful for producing a
CC transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC red coloured fleece. They are useful for producing coloured plant
CC extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC uses include transducing or intensifying an image, providing additional
CC light for growing phototropic organisms e.g. algae and/or corals, for
CC coating materials that experience UV damage e.g. plastics and car
CC upholstery. CFMs are useful in the flower industry, in the development of
CC new varieties of flowering plants. Other contemplated uses include,
CC expression markers, general reporter molecules, photon traps, UV sinks or
CC in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC fungal species, and in fruits and vegetables to enhance their
CC marketability. CFMs embedded in a gel matrix improve image quality in
CC situations of distorted light spectra (biomatrix). The first all-protein
CC chromophore to be isolated was Green Fluorescent protein (GFP). The
CC sequences given in records ABP69924-ABP70048 represent CFM related amino
CC acid sequences. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 13 AA;

Query Match 100.0%; Score 21; DB 5; Length 13;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
|||||
Db 1 SVIAK 5

RESULT 6
ABB99073
ID ABB99073 standard; peptide; 16 AA.
XX
AC ABB99073;
XX
DT 22-JAN-2003 (first entry)
XX
DE N-terminal amino acid sequence of a CFM #13.
XX
KW Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW chromophore; biomatrix; transgenic animal; colouring agent;
KW flower industry; expression marker; reporter molecule; photon trap;
KW UV sink; sunscreen.
XX
OS Unidentified.
XX
PN WO200270703-A2.

```
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 4; Page 281; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 16 AA;

  Query Match             100.0%;  Score 21;  DB 5;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db          1 SVIAK 5


RESULT 7
ABB99074
ID   ABB99074 standard; peptide; 16 AA.
XX
AC   ABB99074;
XX
DT   22-JAN-2003  (first entry)
XX
DE   N-terminal amino acid sequence of a CFM #14.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 10
FT                   /label= Xaa
FT                   /note= "Xaa is any amino acid except Lys"
FT   Misc-difference 11
FT                   /label= Xaa
FT                   /note= "Xaa is any amino acid except Val"
FT   Misc-difference 13
FT                   /label= Xaa
FT                   /note= "Xaa is any amino acid except Met"
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 4; Page 282; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 16 AA;

  Query Match             100.0%;  Score 21;  DB 5;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db          1 SVIAK 5


RESULT 8
```

```
ABB99072
ID   ABB99072 standard; peptide; 16 AA.
XX
AC   ABB99072;
XX
DT   22-JAN-2003  (first entry)
XX
DE   N-terminal amino acid sequence of a CFM #12.
XX
KW   Colour facilitating molecule; CFM; green fluorescent protein; GFP;
KW   chromophore; biomatrix; transgenic animal; colouring agent;
KW   flower industry; expression marker; reporter molecule; photon trap;
KW   UV sink; sunscreen.
XX
OS   Unidentified.
XX
PN   WO200270703-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-GB000928.
XX
PR   02-MAR-2001; 2001US-0273227P.
PR   21-MAR-2001; 2001AU-00003874.
PR   15-OCT-2001; 2001US-0329816P.
XX
PA   (NUFA-) NUFARM LTD.
PA   (UYQU ) UNIV QUEENSLAND.
PA   (JONE/) JONES E L.
XX
PI   Jones EL,  Karan M,  Brugliera F,  Mason J,  Dove SG;
PI   Hoegh-Guldberg IO,  Prescott M;
XX
DR   WPI; 2002-740765/80.
XX
PT   Novel color-facilitating molecule for producing a biomatrix, has a
PT   polypeptide which alone/along with molecules imparts altered visual
PT   characteristics to cells in the absence of excitation by extraneous non-
PT   white light.
XX
PS   Claim 4; Page 281; 510pp; English.
XX
CC   The invention relates to an isolated colour-facilitating molecule (CFM)
CC   comprising a polypeptide which, in a cell, alone or together with one or
CC   more other molecules imparts an altered visual characteristic to the cell
CC   when visualised by a human eye in the absence of excitation by extraneous
CC   non-white light or particle emission. CFMs are useful for producing a
CC   transgenic animal which exhibits a novel colour e.g. sheep with blue or
CC   red coloured fleece. They are useful for producing coloured plant
CC   extracts, e.g. flavouring, beverage or juice or colouring agent. Other
CC   uses include transducing or intensifying an image, providing additional
CC   light for growing phototropic organisms e.g. algae and/or corals, for
CC   coating materials that experience UV damage e.g. plastics and car
CC   upholstery. CFMs are useful in the flower industry, in the development of
CC   new varieties of flowering plants. Other contemplated uses include,
CC   expression markers, general reporter molecules, photon traps, UV sinks or
CC   in sunscreens. CFMs modify visible colour in edible and/or ornamental
CC   fungal species, and in fruits and vegetables to enhance their
CC   marketability. CFMs embedded in a gel matrix improve image quality in
CC   situations of distorted light spectra (biomatrix). The first all-protein
CC   chromophore to be isolated was Green Fluorescent protein (GFP). The
CC   current sequence represents the N-terminal amino acid sequence of a
CC   colour-facilitating molecule (CFM)
XX
SQ   Sequence 16 AA;

  Query Match             100.0%;  Score 21;  DB 5;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db          1 SVIAK 5
```

```
RESULT 9
AAY97148
ID   AAY97148 standard; peptide; 17 AA.
XX
AC   AAY97148;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Pigment protein from coral tissue N-terminal peptide 2.
XX
KW   N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW   tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW   UV filter.
XX
OS   Acropora horrida.
XX
PN   WO200046233-A1.
XX
PD   10-AUG-2000.
XX
PF   02-FEB-2000; 2000WO-AU000056.
XX
PR   02-FEB-1999;   99AU-00008463.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Hoegh-Guldberg O,  Dove S;
XX
DR   WPI; 2000-532892/48.
XX
PT   Novel pigment protein derived from corals capable of emitting
PT   fluorescence upon irradiation by incident light useful as tissue marker,
PT   fluorescent marker or general dyestuff.
XX
PS   Claim 4; Page 42; 49pp; English.
XX
CC   The N-terminal peptides shown in AAY97147-48 are from pigment protein
CC   from coral tissue (PPCT). PPCT is capable of emitting fluorescence upon
CC   irradiation by incident light whose maximal absorbance is in the range of
CC   320-600 nm and a maximal fluorescence emission is in the range of 300-700
CC   nm. PPCT may be used as a tissue marker, fluorescent marker (e.g. to
CC   follow gene expression in transformed tissues) or general dyestuff (all
CC   claimed). PPCT may also be used in sunscreen formulations or UV filters
CC   (both claimed)
XX
SQ   Sequence 17 AA;

  Query Match             100.0%;  Score 21;  DB 3;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db          1 SVIAK 5


RESULT 10
AAY97151
ID   AAY97151 standard; peptide; 18 AA.
XX
AC   AAY97151;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Pigment protein from coral tissue N-terminal peptide 3.
XX
KW   N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW   tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW   UV filter.
XX
OS   Acropora aspera.
```

OS Montipora caliculata.
OS Porites murrayensis.
XX
PN WO200046233-A1.
XX
PD 10-AUG-2000.
XX
PF 02-FEB-2000; 2000WO-AU000056.
XX
PR 02-FEB-1999; 99AU-00008463.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Hoegh-Guldberg O, Dove S;
XX
DR WPI; 2000-532892/48.
XX
PT Novel pigment protein derived from corals capable of emitting
PT fluorescence upon irradiation by incident light useful as tissue marker,
PT fluorescent marker or general dyestuff.
XX
PS Example 2; Page 18; 49pp; English.
XX
CC The N-terminal peptides shown in AAY97151-52 are from pigment protein
CC from coral tissue (PPCT). PPCT is capable of emitting fluorescence upon
CC irradiation by incident light whose maximal absorbance is in the range of
CC 320-600 nm and a maximal fluorescence emission is in the range of 300-700
CC nm. PPCT may be used as a tissue marker, fluorescent marker (e.g. to
CC follow gene expression in transformed tissues) or general dyestuff (all
CC claimed). PPCT may also be used in sunscreen formulations or UV filters
CC (both claimed)
XX
SQ Sequence 18 AA;

Query Match 100.0%; Score 21; DB 3; Length 18;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
|||||
Db 1 SVIAK 5

RESULT 11
AAY97152
ID AAY97152 standard; peptide; 25 AA.
XX
AC AAY97152;
XX
DT 04-DEC-2000 (first entry)
XX
DE Pigment protein from coral tissue N-terminal peptide 4.
XX
KW N-terminal; pigment protein from coral tissue; PPCT; fluorescence;
KW tissue marker; fluorescent marker; dyestuff; sunscreen; ultra violet;
KW UV filter.
XX
OS Porites lobata.
XX
PN WO200046233-A1.
XX
PD 10-AUG-2000.
XX
PF 02-FEB-2000; 2000WO-AU000056.
XX
PR 02-FEB-1999; 99AU-00008463.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Hoegh-Guldberg O, Dove S;
XX
DR WPI; 2000-532892/48.
XX
PT Novel pigment protein derived from corals capable of emitting
PT fluorescence upon irradiation by incident light useful as tissue marker,
PT fluorescent marker or general dyestuff.
XX
PS Example 2; Page 18; 49pp; English.
XX
CC The N-terminal peptides shown in AAY97151-52 are from pigment protein
CC from coral tissue (PPCT). PPCT is capable of emitting fluorescence upon
CC irradiation by incident light whose maximal absorbance is in the range of
CC 320-600 nm and a maximal fluorescence emission is in the range of 300-700
CC nm. PPCT may be used as a tissue marker, fluorescent marker (e.g. to
CC follow gene expression in transformed tissues) or general dyestuff (all
CC claimed). PPCT may also be used in sunscreen formulations or UV filters
CC (both claimed)
XX
SQ Sequence 25 AA;

Query Match 100.0%; Score 21; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
|||||
Db 1 SVIAK 5

RESULT 12
AAB62002
ID AAB62002 standard; protein; 50 AA.
XX
AC AAB62002;
XX
DT 14-MAY-2001 (first entry)
XX
DE C-myb protein fragment (residues 92-141).
XX
KW Cdc5; hCdc5 cell cycle progression; hyperproliferative; cancer; human;
KW coronary artery disease; cellular proliferation; cardiac injury; C-myb;
KW myocardial infarction; cytostatic; cardiant; vasotropic; gene therapy.
XX
OS Homo sapiens.
XX
PN US6183961-B1.
XX
PD 06-FEB-2001.
XX
PF 18-SEP-1998; 98US-00156316.
XX
PR 22-SEP-1997; 97US-0060688P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Bernstein HS, Coughlin SR;
XX
DR WPI; 2001-210295/21.
XX
PT Novel nucleic acids comprising human Cdc5 gene useful in gene therapy
PT techniques for treating diseases related to cell cycle defect such as
PT cancer, coronary artery disease, pulmonary obstructive vascular disease.
XX
PS Disclosure; Fig 2B; 45pp; English.
XX
CC The invention relates to a human Cdc5 (hCdc5) protein. The hCdc5 cDNA is
CC useful in the therapy of pathologies such as diseases, syndromes, or
CC other undesirable conditions resulting from defects in cell cycle
CC progression which may result from hCdc5 gene, in the regulation of the
CC expression of the hCdc5 gene or in a step downstream of hCdc5 in the
CC regulation of cell cycle progression through G2 and entry into mitosis.
CC It is useful for treating a patient having a hyperproliferative disease
CC such as cancer, coronary artery disease, pulmonary vascular obstructive
CC disease, and other disorders of abnormal cellular proliferation (see
CC AAB61997 for various used of the hCdc5 gene). The present sequence
CC represents a fragment of the c-myb protein, used for amino-terminal

```
CC   homology studies with hCdc5 protein
XX
SQ   Sequence 50 AA;

  Query Match             100.0%;  Score 21;  DB 4;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         26 SVIAK 30


RESULT 13
ABG14062
ID   ABG14062 standard; protein; 73 AA.
XX
AC   ABG14062;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #14053.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS78249.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 44421; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 73 AA;

  Query Match             100.0%;  Score 21;  DB 4;  Length 73;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         24 SVIAK 28


RESULT 14
ABP34419
ID   ABP34419 standard; protein; 73 AA.
XX
AC   ABP34419;
XX
DT   09-JUL-2002  (first entry)
XX
DE   Human isomerase-like ORF3392 protein, SEQ ID NO:6784.
XX
KW   Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW   disease monitoring; cytokine; cell proliferation; cell differentiation;
KW   immune modulation; haematopoiesis regulation; tissue growth;
KW   angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW   behaviour; cancer; proliferative disorder; neurological disorder;
KW   cardiovascular disease; immune system disorder; organ transplantation;
KW   tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW   hypothyroidism; cholesterol ester storage disease; infection; vulnerary;
KW   vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
KW   neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW   cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
KW   dermatological; analgesic; virucide; antibacterial; fungicide.
XX
OS   Homo sapiens.
XX
PN   WO200190366-A2.
XX
PD   29-NOV-2001.
XX
PF   24-MAY-2001; 2001WO-US017076.
XX
PR   24-MAY-2000; 2000US-0206690P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Leach MD,  Shimkets RA;
XX
DR   WPI; 2002-106200/14.
DR   N-PSDB; ABN78445.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and disorders related to organ
PT   transplantation.
XX
PS   Claim 10; Page 1943; 2508pp; English.
XX
CC   Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC   designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC   ABN79587 represent cDNAs encoding them. The invention also encompasses
CC   polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC   referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC   the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC   polynucleotides, the recombinant production of ORFX proteins, antibodies
CC   specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC   polypeptides, methods of screening for modulators of ORFX expression or
CC   activity, and methods of screening individuals for a predisposition to an
CC   ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC   range of biological activities, such as cytokine, cell proliferation,
CC   cell differentiation, immune modulation, haematopoiesis regulation,
CC   tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
```

```
CC   chemokinetic activity, haemostatic activity, thrombolytic activity,
CC   receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC   and antiinfective activity, and may also be involved in the determination
CC   of bodily characteristics, fertility and behaviour. ORFX proteins,
CC   nucleic acids and antibodies may be used in the treatment of cancers,
CC   other proliferative disorders such as psoriasis and benign tumours,
CC   neurological disorders such as epilepsy and Alzheimer's disease,
CC   cardiovascular diseases, immune system disorders, disorders related to
CC   organ transplantation, disorders of tissue growth and regeneration,
CC   diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC   storage disease, and infectious diseases caused by viral, bacterial,
CC   fungal and other pathogens. ORFX nucleic acids may also be used as a
CC   source of primers and probes, in the detection of ORFX genomic sequences
CC   or transcripts, in the identification and cloning of homologous
CC   sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC   nucleic acids may additionally be used to produce transgenic animals
CC   which may be useful for studying the function and/or activity of ORFX
CC   protein, and in drug screening. The ORFX proteins may also be used as
CC   immunogens to generate specific antibodies, which are useful in the
CC   diagnosis, treatment and monitoring of ORFX-associated diseases
XX
SQ   Sequence 73 AA;

  Query Match             100.0%;  Score 21;  DB 5;  Length 73;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         64 SVIAK 68


RESULT 15
ABP77483
ID   ABP77483 standard; protein; 75 AA.
XX
AC   ABP77483;
XX
DT   07-MAR-2003  (first entry)
XX
DE   N. gonorrhoeae amino acid sequence SEQ ID 1496.
XX
KW   Antibacterial; infection; vaccine; gene therapy.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO200279243-A2.
XX
PD   10-OCT-2002.
XX
PF   12-FEB-2002; 2002WO-IB002069.
XX
PR   12-FEB-2001; 2001GB-00003424.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Fontana MR,  Pizza M,  Masignani V,  Monaci E;
XX
DR   WPI; 2003-058415/05.
DR   N-PSDB; ABZ38453.
XX
PT   New protein from Neisseria gonorrheae, useful for the manufacture of a
PT   medicament for treating or preventing N. gonorrheae infection.
XX
PS   Disclosure; Page 296; 815pp; English.
XX
CC   The present invention relates to proteins from Neisseria gonorrhoeae.
CC   Also disclosed are the nucleic acid molecules encoding the proteins and
CC   antibodies that specifically bind to the proteins. The composition
CC   comprising the protein, nucleic acid or antibody is useful for the
CC   manufacture of a medicament for treating or preventing N. gonorrhoeae
CC   infection, this may be in the form of a vaccine or gene therapy.
CC   Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC   molecules of the invention
XX
SQ   Sequence 75 AA;

  Query Match             100.0%;  Score 21;  DB 6;  Length 75;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         30 SVIAK 34


Search completed: August 12, 2004, 06:17:04
Job time : 3.89549 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:19:43 ; Search time 10.5225 Seconds
(without alignments)
149.169 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 1292805

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	100.0	50	9	US-09-757-049A-6	Sequence 6, Appli
2	21	100.0	52	9	US-09-912-962-14	Sequence 14, Appl
3	21	100.0	55	12	US-10-424-599-235707	Sequence 235707,
4	21	100.0	63	16	US-10-437-963-201620	Sequence 201620,
5	21	100.0	66	12	US-10-424-599-281604	Sequence 281604,
6	21	100.0	72	12	US-10-424-599-259733	Sequence 259733,
7	21	100.0	73	11	US-09-864-408A-6784	Sequence 6784, Ap
8	21	100.0	76	12	US-10-276-774-2542	Sequence 2542, Ap
9	21	100.0	102	9	US-09-738-973-196	Sequence 196, App
10	21	100.0	102	9	US-09-854-133-196	Sequence 196, App
11	21	100.0	102	14	US-10-144-649A-196	Sequence 196, App
12	21	100.0	107	12	US-10-424-599-200366	Sequence 200366,
13	21	100.0	138	12	US-10-424-599-252896	Sequence 252896,
14	21	100.0	142	12	US-10-425-114-39108	Sequence 39108, A
15	21	100.0	156	10	US-09-892-398-4	Sequence 4, Appli
16	21	100.0	156	10	US-09-892-398-46	Sequence 46, Appl
17	21	100.0	161	12	US-10-424-599-283302	Sequence 283302,
18	21	100.0	165	12	US-10-424-599-280513	Sequence 280513,
19	21	100.0	167	15	US-10-369-493-10328	Sequence 10328, A
20	21	100.0	169	16	US-10-437-963-137115	Sequence 137115,
21	21	100.0	189	12	US-10-424-599-236262	Sequence 236262,
22	21	100.0	192	12	US-10-425-114-64027	Sequence 64027, A
23	21	100.0	193	12	US-10-424-599-180029	Sequence 180029,
24	21	100.0	196	12	US-10-425-114-72020	Sequence 72020, A
25	21	100.0	196	12	US-10-425-114-72021	Sequence 72021, A
26	21	100.0	208	12	US-10-424-599-209489	Sequence 209489,
27	21	100.0	233	14	US-10-156-761-9990	Sequence 9990, Ap
28	21	100.0	242	12	US-10-424-599-252176	Sequence 252176,
29	21	100.0	244	15	US-10-369-493-21085	Sequence 21085, A
30	21	100.0	249	14	US-10-156-275-88	Sequence 88, Appl
31	21	100.0	260	10	US-09-935-338-105	Sequence 105, App
32	21	100.0	260	12	US-10-380-430-9	Sequence 9, Appli
33	21	100.0	263	12	US-10-282-122A-64121	Sequence 64121, A
34	21	100.0	268	12	US-10-282-122A-63475	Sequence 63475, A
35	21	100.0	273	16	US-10-287-226-374	Sequence 374, App
36	21	100.0	283	12	US-10-425-114-53397	Sequence 53397, A
37	21	100.0	283	12	US-10-425-114-71886	Sequence 71886, A
38	21	100.0	283	12	US-10-425-114-72284	Sequence 72284, A
39	21	100.0	319	9	US-09-864-761-34819	Sequence 34819, A
40	21	100.0	320	9	US-09-886-055-121	Sequence 121, App
41	21	100.0	320	10	US-09-804-291-121	Sequence 121, App
42	21	100.0	320	11	US-09-844-861A-6	Sequence 6, Appli
43	21	100.0	320	12	US-10-343-650A-206	Sequence 206, App
44	21	100.0	320	14	US-10-017-161-942	Sequence 942, App
45	21	100.0	320	15	US-10-292-798-816	Sequence 816, App

ALIGNMENTS

RESULT 1
US-09-757-049A-6
; Sequence 6, Application US/09757049A
; Patent No. US20020127702A1
; GENERAL INFORMATION:
; APPLICANT: BERNSTEIN, Harold S.
; APPLICANT: COUGHLIN, Shaun R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR REGULATING CELL CYCLE
; TITLE OF INVENTION: PROGRESSION
; FILE REFERENCE: UCSF-020/02US
; CURRENT APPLICATION NUMBER: US/09/757,049A
; CURRENT FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/156,316
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: US 60/060,688
; PRIOR FILING DATE: 1997-09-22
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-757-049A-6

Query Match 100.0%; Score 21; DB 9; Length 50;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 SVIAK 5
              |||||
Db         26 SVIAK 30
```

RESULT 2
US-09-912-962-14
; Sequence 14, Application US/09912962
; Patent No. US20020076719A1
; GENERAL INFORMATION:

```
;       APPLICANT: de Lange, Titia
;                  Broccoli, Dominique
;                  Smogorzewska, Agata
;       TITLE OF INVENTION: TELOMERE REPEAT BINDING FACTOR AND
;                           DIAGNOSTIC AND THERAPEUTIC USE THEREOF
;       NUMBER OF SEQUENCES: 52
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: KLAUBER & JACKSON
;            STREET: 411 Hackensack Avenue
;            CITY: Hackensack
;            STATE: New Jersey
;            COUNTRY: USA
;            ZIP: 07601
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/912,962
;            FILING DATE: 25-Jul-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/018,635
;            FILING DATE: 04-FEB-1998
;       ATTORNEY/AGENT INFORMATION:
;            NAME: David A. Jackson
;            REGISTRATION NUMBER: 26,742
;            REFERENCE/DOCKET NUMBER: 600-1-142 CIP1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 201-487-5800
;            TELEFAX: 201-343-1684
;            TELEX: 133521
;   INFORMATION FOR SEQ ID NO: 14:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 52 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-912-962-14

  Query Match             100.0%;  Score 21;  DB 9;  Length 52;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         27 SVIAK 31


RESULT 3
US-10-424-599-235707
; Sequence 235707, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 235707
;   LENGTH: 55
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_54872C.1.pep
US-10-424-599-235707

  Query Match             100.0%;  Score 21;  DB 12;  Length 55;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         25 SVIAK 29


RESULT 4
US-10-437-963-201620
; Sequence 201620, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 201620
;   LENGTH: 63
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_96978C.1.pep
US-10-437-963-201620

  Query Match             100.0%;  Score 21;  DB 16;  Length 63;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         22 SVIAK 26


RESULT 5
US-10-424-599-281604
; Sequence 281604, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 281604
;   LENGTH: 66
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_96309C.1.pep
US-10-424-599-281604

  Query Match             100.0%;  Score 21;  DB 12;  Length 66;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 SVIAK 5
              |||||
Db         57 SVIAK 61


RESULT 6
US-10-424-599-259733
; Sequence 259733, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 259733
;   LENGTH: 72
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_76564C.1.pep
US-10-424-599-259733

  Query Match             100.0%;  Score 21;  DB 12;  Length 72;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         32 SVIAK 36


RESULT 7
US-09-864-408A-6784
; Sequence 6784, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
;  APPLICANT: Leach, Martin D.
;  APPLICANT:  Shimkets, Richard A.
;  TITLE OF INVENTION: No. US20040009474A1el Human Polynucleotides and Polypeptides Encc
;  FILE REFERENCE: 21402-012
;  CURRENT APPLICATION NUMBER: US/09/864,408A
;  CURRENT FILING DATE:  2001-05-24
;  PRIOR APPLICATION NUMBER: 60/206,690
;  PRIOR FILING DATE: 2000-05-24
;  NUMBER OF SEQ ID NOS: 9068
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6784
;   LENGTH: 73
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (17)..(17)
;   OTHER INFORMATION: Wherein Xaa may be any naturally occuring amino acid
US-09-864-408A-6784

  Query Match             100.0%;  Score 21;  DB 11;  Length 73;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         64 SVIAK 68


RESULT 8
US-10-276-774-2542
; Sequence 2542, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  APPLICANT:  Tang, Y, Tom et al
;  TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
;  FILE REFERENCE: 21272-030
;  CURRENT APPLICATION NUMBER: US/10/276,774
;  CURRENT FILING DATE:  2002-11-18
;  PRIOR APPLICATION NUMBER: 09/560,875
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 09/496,914
;  PRIOR FILING DATE: 2000-02-03
;  NUMBER OF SEQ ID NOS: 2700
;  SOFTWARE: Custom
; SEQ ID NO 2542
;   LENGTH: 76
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(76)
;   OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-276-774-2542

  Query Match             100.0%;  Score 21;  DB 12;  Length 76;
  Best Local Similarity   100.0%;  Pred. No. 3.2e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         24 SVIAK 28


RESULT 9
US-09-738-973-196
; Sequence 196, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Fling, Steven P.
;  APPLICANT:  Mohamath, Raodoh
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Indirias, Carol Yoseph
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Elliot, Mark
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Kalos, Michael D.
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  THE THERAPY AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.475C9
;  CURRENT APPLICATION NUMBER: US/09/738,973
;  CURRENT FILING DATE:  2000-12-14
;  NUMBER OF SEQ ID NOS:  587
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 196
;   LENGTH: 102
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-738-973-196

  Query Match             100.0%;  Score 21;  DB 9;  Length 102;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         44 SVIAK 48
```

```
RESULT 10
US-09-854-133-196
; Sequence 196, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
;  APPLICANT: Lodes, Michael J.
;  APPLICANT:  Mohamath, Raodoh
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  THE THERAPY AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.475C10
;  CURRENT APPLICATION NUMBER: US/09/854,133
;  CURRENT FILING DATE:  2001-05-11
;  NUMBER OF SEQ ID NOS: 735
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 196
;   LENGTH: 102
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-854-133-196

  Query Match             100.0%;  Score 21;  DB 9;  Length 102;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         44 SVIAK 48


RESULT 11
US-10-144-649A-196
; Sequence 196, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
;  APPLICANT: Lodes, Michael J.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  McNeill, Patricia D.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  THE THERAPY AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.475C11
;  CURRENT APPLICATION NUMBER: US/10/144,649A
;  CURRENT FILING DATE:  2002-08-21
;  NUMBER OF SEQ ID NOS: 749
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 196
;   LENGTH: 102
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-10-144-649A-196

  Query Match             100.0%;  Score 21;  DB 14;  Length 102;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         44 SVIAK 48


RESULT 12
US-10-424-599-200366
; Sequence 200366, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 200366
;   LENGTH: 107
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(107)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_22957C.1.pep
US-10-424-599-200366

  Query Match             100.0%;  Score 21;  DB 12;  Length 107;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db          5 SVIAK 9


RESULT 13
US-10-424-599-252896
; Sequence 252896, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 252896
;   LENGTH: 138
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_70391C.1.pep
US-10-424-599-252896

  Query Match             100.0%;  Score 21;  DB 12;  Length 138;
  Best Local Similarity   100.0%;  Pred. No. 6e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db        125 SVIAK 129


RESULT 14
US-10-425-114-39108
; Sequence 39108, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
```

```
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 39108
;   LENGTH: 142
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 700208712_FLI.pep
US-10-425-114-39108

  Query Match             100.0%;  Score 21;  DB 12;  Length 142;
  Best Local Similarity   100.0%;  Pred. No. 6.2e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         41 SVIAK 45


RESULT 15
US-09-892-398-4
; Sequence 4, Application US/09892398
; Publication No. US20030028002A1
;   GENERAL INFORMATION:
;        APPLICANT: Hirai, Hiroshi
;                   Sherr, Charles
;                   Inoue, Kazushi
;                   Bodner, Sarah M.
;        TITLE OF INVENTION: CYCLIN-D BINDING FACTOR, AND USES
;                            THEREOF
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: David A. Jackson, Esq.
;             STREET: 411 Hackensack Ave, Continental Plaza, 4th
;                     Floor
;             CITY: Hackensack
;             STATE: New Jersey
;             COUNTRY: USA
;             ZIP: 07601
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/892,398
;             FILING DATE: 27-Jun-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/280,590
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Jackson Esq., David A.
;             REGISTRATION NUMBER: 26,742
;             REFERENCE/DOCKET NUMBER: 1340-1-002 N CP2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 201-487-5800
;             TELEFAX: 201-343-1684
;   INFORMATION FOR SEQ ID NO: 4:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 156 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: NO
;        FRAGMENT TYPE: internal
;        ORIGINAL SOURCE:
;             ORGANISM: Mus musculus
;        SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-892-398-4

  Query Match             100.0%;  Score 21;  DB 10;  Length 156;
  Best Local Similarity   100.0%;  Pred. No. 6.8e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         79 SVIAK 83


Search completed: August 12, 2004, 06:51:20
Job time : 10.5225 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 1.30123 Seconds
(without alignments)
1212.385 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	100.0	63	6	Q9XT66	Q9xt66 canis famil
2	21	100.0	75	16	Q8CRL6	Q8crl6 staphylococ
3	21	100.0	95	16	Q8EVW5	Q8evw5 mycoplasma
4	21	100.0	98	16	Q8P4C6	Q8p4c6 xanthomonas
5	21	100.0	107	5	Q7YUB8	Q7yub8 aphelenchus
6	21	100.0	111	11	Q922Z7	Q922z7 mus musculu
7	21	100.0	126	9	Q8HAJ2	Q8haj2 bacteriopha
8	21	100.0	126	11	Q8C8J3	Q8c8j3 mus musculu
9	21	100.0	132	5	O17211	O17211 caenorhabdi
10	21	100.0	135	9	Q8SC48	Q8sc48 stx2 conver
11	21	100.0	141	16	Q8E7B3	Q8e7b3 streptococc
12	21	100.0	141	16	Q89P48	Q89p48 bradyrhizob
13	21	100.0	152	16	Q8VK36	Q8vk36 mycobacteri
14	21	100.0	152	16	Q7U080	Q7u080 mycobacteri
15	21	100.0	161	9	Q9XJQ6	Q9xjq6 bacteriopha
16	21	100.0	161	17	Q8ZVL5	Q8zvl5 pyrobaculum
17	21	100.0	162	9	Q8HA15	Q8ha15 bacteriopha
18	21	100.0	164	16	Q8EXA9	Q8exa9 leptospira
19	21	100.0	171	16	Q98R43	Q98r43 mycoplasma
20	21	100.0	172	9	Q8HAE9	Q8hae9 salmonella
21	21	100.0	175	3	Q00300	Q00300 ajellomyces
22	21	100.0	175	3	Q9P436	Q9p436 ajellomyces
23	21	100.0	175	3	Q9P433	Q9p433 ajellomyces
24	21	100.0	175	3	Q9P439	Q9p439 ajellomyces
25	21	100.0	175	3	Q9P435	Q9p435 ajellomyces
26	21	100.0	175	3	Q9P437	Q9p437 ajellomyces
27	21	100.0	175	3	Q9P434	Q9p434 ajellomyces
28	21	100.0	176	13	P87467	P87467 gallus gall
29	21	100.0	177	11	Q9CVJ7	Q9cvj7 mus musculu
30	21	100.0	181	16	Q8YY51	Q8yy51 anabaena sp
31	21	100.0	188	16	Q99V25	Q99v25 staphylococ
32	21	100.0	188	16	Q8NX89	Q8nx89 staphylococ
33	21	100.0	191	11	P97753	P97753 mus sp. gag
34	21	100.0	192	16	Q8R852	Q8r852 thermoanaer
35	21	100.0	196	16	Q8G813	Q8g813 bifidobacte
36	21	100.0	197	16	Q9AAV8	Q9aav8 caulobacter
37	21	100.0	199	2	O52946	O52946 bacillus su
38	21	100.0	200	16	Q8Y9M4	Q8y9m4 listeria mo
39	21	100.0	203	10	Q9SW28	Q9sw28 arabidopsis
40	21	100.0	209	16	Q8FSI1	Q8fsi1 corynebacte
41	21	100.0	218	5	Q9N6J5	Q9n6j5 leishmania
42	21	100.0	219	16	Q9F8V9	Q9f8v9 agrobacteri
43	21	100.0	221	5	Q95P04	Q95p04 goniopora t
44	21	100.0	223	2	Q9AHZ2	Q9ahz2 photorhabdu
45	21	100.0	224	12	Q69112	Q69112 herpes simp

ALIGNMENTS

```
RESULT 1
Q9XT66
ID   Q9XT66       PRELIMINARY;      PRT;    63 AA.
AC   Q9XT66;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   C-myb (Fragment).
GN   C-MYB.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99265967; PubMed=10331940;
RA   Li R., Mignot E., Faraco J., Kadotani H., Cantanese J., Zhao B.,
RA   Lin X., Hinton L., Ostrander E.A., Patterson D.F., de Jong P.J.;
RT   "Construction and characterization of an eightfold redundant dog
RT   genomic bacterial artificial chromosome library.";
RL   Genomics 58:9-17(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kodatani H., Mignot E.;
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
DR   EMBL; AF103748; AAD40574.1; -.
DR   HSSP; P06876; 1MBG.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   Pfam; PF00249; myb_DNA-binding; 2.
DR   SMART; SM00717; SANT; 1.
DR   PROSITE; PS00037; MYB_1; 1.
DR   PROSITE; PS00334; MYB_2; 1.
DR   PROSITE; PS50090; MYB_3; 1.
KW   Nuclear protein.
FT   NON_TER       1       1
```

```
FT   NON_TER      63     63
SQ   SEQUENCE   63 AA;  7707 MW;  D8C86265802F3C9F CRC64;

  Query Match             100.0%;  Score 21;  DB 6;  Length 63;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         19 SVIAK 23


RESULT 2
Q8CRL6
ID   Q8CRL6      PRELIMINARY;      PRT;    75 AA.
AC   Q8CRL6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Conserved hypothetical protein.
GN   SE1742.
OS   Staphylococcus epidermidis.
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=1282;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 12228;
RA   Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA   Chen Z., Wen Y.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE016749; AAO05341.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   75 AA;  8090 MW;  9B017C60D9C61D9F CRC64;

  Query Match             100.0%;  Score 21;  DB 16;  Length 75;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         71 SVIAK 75


RESULT 3
Q8EVW5
ID   Q8EVW5      PRELIMINARY;      PRT;    95 AA.
AC   Q8EVW5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Ribosomal protein L27.
GN   MYPE4440.
OS   Mycoplasma penetrans.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=28227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HF-2;
RX   MEDLINE=22354719; PubMed=12466555;
RA   Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA   Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT   "The complete genomic sequence of Mycoplasma penetrans, an
RT   intracellular bacterial pathogen in humans.";
RL   Nucleic Acids Res. 30:5293-5300(2002).
DR   EMBL; AP004171; BAC44234.1; -.
DR   GO; GO:0005622; C:intracellular; IEA.
DR   GO; GO:0005840; C:ribosome; IEA.
DR   GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR   GO; GO:0006412; P:protein biosynthesis; IEA.
DR   InterPro; IPR001684; Ribosomal_L27.
DR   Pfam; PF01016; Ribosomal_L27; 1.
DR   PRINTS; PR00063; RIBOSOMALL27.
DR   ProDom; PD003114; Ribosomal_L27; 1.
DR   TIGRFAMs; TIGR00062; L27; 1.
DR   PROSITE; PS00831; RIBOSOMAL_L27; 1.
KW   Complete proteome.
SQ   SEQUENCE   95 AA;  10464 MW;  735D951C94B7A730 CRC64;

  Query Match             100.0%;  Score 21;  DB 16;  Length 95;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         91 SVIAK 95


RESULT 4
Q8P4C6
ID   Q8P4C6      PRELIMINARY;      PRT;    98 AA.
AC   Q8P4C6;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein XCC3784.
GN   XCC3784.
OS   Xanthomonas campestris (pv. campestris).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=340;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33913 / NCPPB 528;
RX   MEDLINE=22022145; PubMed=12024217;
RA   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA   Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA   Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA   Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA   Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA   Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA   Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA   Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA   Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA   Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA   Setubal J.C., Kitajima J.P.;
RT   "Comparison of the genomes of two Xanthomonas pathogens with differing
RT   host specificities.";
RL   Nature 417:459-463(2002).
DR   EMBL; AE012499; AAM43030.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0016998; P:cell wall catabolism; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR002482; LysM.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF01476; LysM; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   98 AA;  10593 MW;  0DB6218EB6AAFA60 CRC64;

  Query Match             100.0%;  Score 21;  DB 16;  Length 98;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         56 SVIAK 60


RESULT 5
Q7YUB8
```

```
ID   Q7YUB8      PRELIMINARY;      PRT;   107 AA.
AC   Q7YUB8;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Glutaredoxin.
GN   GLX-1.
OS   Aphelenchus avenae (Mycophagous nematode).
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Aphelenchoidea;
OC   Aphelenchidae; Aphelenchus.
OX   NCBI_TaxID=70226;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Browne J.A., Goyal K., Tunnacliffe A., Burnell A.;
RT   "Expression of a glutaredoxin gene induced by desiccation and
RT   oxidative stress in the anhydrobiotic nematode Aphelehus avenae.";
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY340999; AAQ20895.1; -.
SQ   SEQUENCE   107 AA;  11614 MW;  CB0396A67FEC9C32 CRC64;

  Query Match             100.0%;  Score 21;  DB 5;  Length 107;
  Best Local Similarity   100.0%;  Pred. No. 3.3e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         10 SVIAK 14


RESULT 6
Q922Z7
ID   Q922Z7      PRELIMINARY;      PRT;   111 AA.
AC   Q922Z7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to DNA polymerase beta.
GN   POLB.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC006681; AAH06681.1; -.
DR   MGD; MGI:97740; Polb.
DR   GO; GO:0006916; P:anti-apoptosis; IMP.
DR   GO; GO:0008220; P:necrosis; IMP.
DR   InterPro; IPR003583; HHH_1.
DR   SMART; SM00278; HhH1; 1.
SQ   SEQUENCE   111 AA;  12247 MW;  E81BBACDF1A3B44F CRC64;

  Query Match             100.0%;  Score 21;  DB 11;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 3.4e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         44 SVIAK 48


RESULT 7
Q8HAJ2
ID   Q8HAJ2      PRELIMINARY;      PRT;   126 AA.
AC   Q8HAJ2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Antitermination protein Q.
GN   Q.
OS   Bacteriophage LC159.
OC   Viruses.
OX   NCBI_TaxID=210928;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=159;
RA   Muniesa M., Jofre J.;
RT   "Variability of shiga converting bacteriophages in E. coli O157:H7
RT   strains of human origin isolated from the same outbreak.";
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF548456; AAN59919.1; -.
SQ   SEQUENCE   126 AA;  14230 MW;  B8F1776A0329F55A CRC64;

  Query Match             100.0%;  Score 21;  DB 9;  Length 126;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db          3 SVIAK 7


RESULT 8
Q8C8J3
ID   Q8C8J3      PRELIMINARY;      PRT;   126 AA.
AC   Q8C8J3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical S-adenosyl-L-methionine-dependent methyltransferases
DE   structure containing protein.
GN   4732479N06RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK046928; BAC32921.1; -.
DR   MGD; MGI:2442530; 4732479N06Rik.
KW   Hypothetical protein.
SQ   SEQUENCE   126 AA;  14568 MW;  0AB92E67189578CD CRC64;

  Query Match             100.0%;  Score 21;  DB 11;  Length 126;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         45 SVIAK 49


RESULT 9
O17211
ID   O17211      PRELIMINARY;      PRT;   132 AA.
AC   O17211;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   C01B12.7 protein.
GN   C01B12.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=Bristol N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Scheet P., Maggi L.;
RT   "The sequence of C. elegans cosmid C01B12.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF025458; AAB70973.1; -.
DR   PIR; T32373; T32373.
DR   WormPep; C01B12.7; CE07795.
SQ   SEQUENCE   132 AA;  15750 MW;  A2C8BA7465940DF2 CRC64;

  Query Match             100.0%;  Score 21;  DB 5;  Length 132;
  Best Local Similarity   100.0%;  Pred. No. 4e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         25 SVIAK 29


RESULT 10
Q8SC48
ID   Q8SC48      PRELIMINARY;      PRT;   135 AA.
AC   Q8SC48;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Stx2 converting bacteriophage I.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC   Lambda-like viruses.
OX   NCBI_TaxID=180816;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Stx2 phage-I;
RA   Sato T., Shimizu T., Watarai M., Kobayashi M., Kano S., Hamabata T.,
RA   Yamasaki S., Takeda Y.;
RT   "Genomic sequence of Shiga toxin 2-converting phage isolated from
RT   Escherichia coli O157:H7 Okayama strain and comparison with other
RT   Shiga toxin 2-converting phages.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP004402; BAB87967.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR000531; TonB_boxC.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   135 AA;  16106 MW;  15A614C2A739178C CRC64;

  Query Match             100.0%;  Score 21;  DB 9;  Length 135;
  Best Local Similarity   100.0%;  Pred. No. 4.1e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db        100 SVIAK 104


RESULT 11
Q8E7B3
ID   Q8E7B3      PRELIMINARY;      PRT;   141 AA.
AC   Q8E7B3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   GBS0242.
OS   Streptococcus agalactiae (serotype III).
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=216495;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NEM316 / Serotype III;
RX   MEDLINE=22242508; PubMed=12354221;
RA   Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA   Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA   Kunst F.;
RT   "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT   invasive neonatal disease.";
RL   Mol. Microbiol. 45:1499-1513(2002).
DR   EMBL; AL766844; CAD45887.1; -.
DR   SagaList; gbs0242; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   141 AA;  15937 MW;  924D2E86930763F5 CRC64;

  Query Match             100.0%;  Score 21;  DB 16;  Length 141;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         21 SVIAK 25


RESULT 12
Q89P48
ID   Q89P48      PRELIMINARY;      PRT;   141 AA.
AC   Q89P48;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Bll3635 protein.
GN   BLL3635.
OS   Bradyrhizobium japonicum.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bradyrhizobiaceae; Bradyrhizobium.
OX   NCBI_TaxID=375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=USDA 110;
RX   MEDLINE=22484998; PubMed=12597275;
RA   Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA   Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA   Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA   Tabata S.;
RT   "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT   Bradyrhizobium japonicum USDA110.";
RL   DNA Res. 9:189-197(2002).
DR   EMBL; AP005948; BAC48900.1; -.
KW   Complete proteome.
SQ   SEQUENCE   141 AA;  14457 MW;  9F10019F39AD214B CRC64;

  Query Match             100.0%;  Score 21;  DB 16;  Length 141;
```

```
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         56 SVIAK 60


RESULT 13
Q8VK36
ID   Q8VK36      PRELIMINARY;      PRT;   152 AA.
AC   Q8VK36;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   4-hydroxyphenylpyruvate dioxygenase C terminal domain containing
DE   protein.
GN   MT1364.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE007009; AAK45627.1; -.
DR   TIGR; MT1364; -.
DR   GO; GO:0016702; F:oxidoreductase activity, acting on single d. . .; IEA.
DR   InterPro; IPR004360; Gly_bleo_diox.
DR   Pfam; PF00903; Glyoxalase; 1.
KW   Dioxygenase; Pyruvate.
SQ   SEQUENCE   152 AA;  16626 MW;  60E64662DC2B343D CRC64;

  Query Match             100.0%;  Score 21;  DB 16;  Length 152;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         81 SVIAK 85


RESULT 14
Q7U080
ID   Q7U080      PRELIMINARY;      PRT;   152 AA.
AC   Q7U080;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Conserved hypothetical protein.
GN   MB1357C.
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1765;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AF2122/97;
RX   MEDLINE=22709107; PubMed=12788972;
RA   Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA   Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA   Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA   Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT   "The complete genome sequence of Mycobacterium bovis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR   EMBL; BX248338; CAD94218.1; -.
KW   Complete proteome.
SQ   SEQUENCE   152 AA;  16626 MW;  60E64662DC2B343D CRC64;

  Query Match             100.0%;  Score 21;  DB 16;  Length 152;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         81 SVIAK 85


RESULT 15
Q9XJQ6
ID   Q9XJQ6      PRELIMINARY;      PRT;   161 AA.
AC   Q9XJQ6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Q protein.
GN   Q.
OS   Bacteriophage 21.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX   NCBI_TaxID=10743;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20092464; PubMed=10628842;
RA   Karch H., Schmidt H., Janetzki-Mittmann C., Scheef J., Kroeger M.;
RT   "Shiga toxins even when different are encoded at identical positions
RT   in the genomes of related temperate bacteriophages.";
RL   Mol. Gen. Genet. 262:600-607(1999).
DR   EMBL; AJ237660; CAB39993.1; -.
SQ   SEQUENCE   161 AA;  18497 MW;  A1124675BE0F5896 CRC64;

  Query Match             100.0%;  Score 21;  DB 9;  Length 161;
  Best Local Similarity   100.0%;  Pred. No. 4.9e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         37 SVIAK 41


Search completed: August 12, 2004, 06:19:32
Job time : 13.5512 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 0.307377 Seconds
(without alignments)
847.008 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	100.0	86	1	MINE_AGRT5	Q8uax0 agrobacteri
2	21	100.0	95	1	RL27_MYCPE	Q8evw5 mycoplasma
3	21	100.0	101	1	SR19_METAC	Q8tty3 methanosarc
4	21	100.0	101	1	SR19_METMA	Q8pwm7 methanosarc
5	21	100.0	134	1	YA7B_HAEIN	Q57425 haemophilus
6	21	100.0	188	1	PUR3_STAAM	Q99v25 staphylococ
7	21	100.0	188	1	PUR3_STAAW	Q8nx89 staphylococ
8	21	100.0	224	1	RNH2_SYNPX	Q7u4c6 synechococc
9	21	100.0	224	1	VGLL_HSV2H	P28278 herpes simp
10	21	100.0	225	1	RS3_THEAC	Q9hir5 thermoplasm
11	21	100.0	225	1	RS3_THEVO	Q97bx1 thermoplasm
12	21	100.0	233	1	RNH2_STRAW	Q82kf0 streptomyce
13	21	100.0	236	1	FLGD_BUCAI	P57421 buchnera ap
14	21	100.0	250	1	VNST_PTPV	P03516 punta toro
15	21	100.0	253	1	EXBB_XANCP	O34260 xanthomonas
16	21	100.0	263	1	KSGA_MYCPN	P75113 m dimethyla
17	21	100.0	267	1	RS3_MYCGE	P47403 mycoplasma
18	21	100.0	271	1	EL2_RAT	P00774 rattus norv
19	21	100.0	282	1	CY1_RHOVI	P81379 rhodopseudo
20	21	100.0	293	1	MATA_NEUCR	P19392 neurospora
21	21	100.0	320	1	OYN1_HUMAN	Q8nh53 homo sapien
22	21	100.0	331	1	Y542_RICCN	Q92i78 rickettsia
23	21	100.0	333	1	DPOB_XENLA	O57383 xenopus lae
24	21	100.0	334	1	DPOB_HUMAN	P06746 homo sapien
25	21	100.0	334	1	DPOB_RAT	P06766 rattus norv
26	21	100.0	342	1	AQP7_HUMAN	O14520 homo sapien
27	21	100.0	415	1	MUA1_THETN	Q8rd88 thermoanaer
28	21	100.0	430	1	RFBX_SALTY	P26400 salmonella
29	21	100.0	474	1	SAHH_RALSO	Q8y387 ralstonia s
30	21	100.0	502	1	TBG_CANAL	O93807 candida alb
31	21	100.0	528	1	CTK1_YEAST	Q03957 saccharomyc
32	21	100.0	562	1	SYR_LACPL	Q88x53 lactobacill
33	21	100.0	563	1	SYR_ENTFA	Q831n1 enterococcu
34	21	100.0	564	1	SYR_LACLA	Q9ce12 lactococcus
35	21	100.0	572	1	YHM4_YEAST	P38788 saccharomyc
36	21	100.0	593	1	KPYA_TOBAC	Q40545 nicotiana t
37	21	100.0	602	1	HOXF_ALCEU	P22317 alcaligenes
38	21	100.0	617	1	CHIT_CAEEL	Q11174 caenorhabdi
39	21	100.0	624	1	MYB_XENLA	Q08759 xenopus lae
40	21	100.0	633	1	CLPX_HUMAN	O76031 homo sapien
41	21	100.0	634	1	CLPX_MOUSE	Q9jhs4 mus musculu
42	21	100.0	636	1	MYB_MOUSE	P06876 mus musculu
43	21	100.0	640	1	MYB_BOVIN	P46200 bos taurus
44	21	100.0	640	1	MYB_HUMAN	P10242 homo sapien
45	21	100.0	641	1	MYB_CHICK	P01103 gallus gall

ALIGNMENTS

RESULT 1
MINE_AGRT5
ID MINE_AGRT5 STANDARD; PRT; 86 AA.
AC Q8UAX0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cell division topological specificity factor.
GN MINE OR ATU3247 OR AGR_L_3134.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
CC -!- FUNCTION: Prevents the cell division inhibition by proteins minC
CC and minD at internal division sites while permitting inhibition at
CC polar sites. This ensures cell division at the proper site by
CC restricting the formation of a division septum at the midpoint of
CC the long axis of the cell (By similarity).
CC -!- SIMILARITY: Belongs to the minE family.
CC --

CC --

```
DR   EMBL; AE009254; AAL44063.1; -.
DR   EMBL; AE008359; AAK90142.1; -.
DR   PIR; AI2955; AI2955.
DR   PIR; D98327; D98327.
DR   HAMAP; MF_00262; -; 1.
DR   InterPro; IPR005527; MinE.
DR   Pfam; PF03776; MinE; 1.
DR   TIGRFAMs; TIGR01215; minE; 1.
KW   Cell division; Complete proteome.
SQ   SEQUENCE   86 AA;  9703 MW;  B0E274F6A48D52F2 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 86;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         45 SVIAK 49


RESULT 2
RL27_MYCPE
ID   RL27_MYCPE     STANDARD;      PRT;    95 AA.
AC   Q8EVW5;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   50S ribosomal protein L27.
GN   RPMA OR MYPE4440.
OS   Mycoplasma penetrans.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=28227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HF-2;
RX   MEDLINE=22354719; PubMed=12466555;
RA   Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA   Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT   "The complete genomic sequence of Mycoplasma penetrans, an
RT   intracellular bacterial pathogen in humans.";
RL   Nucleic Acids Res. 30:5293-5300(2002).
CC   -!- SIMILARITY: Belongs to the L27P family of ribosomal proteins.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AP004171; BAC44234.1; -.
DR   HAMAP; MF_00539; -; 1.
DR   InterPro; IPR001684; Ribosomal_L27.
DR   Pfam; PF01016; Ribosomal_L27; 1.
DR   PRINTS; PR00063; RIBOSOMALL27.
DR   ProDom; PD003114; Ribosomal_L27; 1.
DR   TIGRFAMs; TIGR00062; L27; 1.
DR   PROSITE; PS00831; RIBOSOMAL_L27; 1.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   95 AA;  10464 MW;  735D951C94B7A730 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 95;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         91 SVIAK 95


RESULT 3
SR19_METAC
ID   SR19_METAC     STANDARD;      PRT;   101 AA.
AC   Q8TTY3;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Signal recognition particle 19 kDa protein (SRP19).
GN   SRP19 OR MA0292.
OS   Methanosarcina acetivorans.
OC   Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC   Methanosarcinaceae; Methanosarcina.
OX   NCBI_TaxID=2214;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C2A / ATCC 35395 / DSM 2834;
RX   MEDLINE=21929760; PubMed=11932238;
RA   Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA   FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA   Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA   Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA   Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA   Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA   Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA   Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA   Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA   Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA   Metcalf W.W., Birren B.;
RT   "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT   and physiological diversity.";
RL   Genome Res. 12:532-542(2002).
CC   -!- FUNCTION: Signal-recognition-particle assembly, binds directly to
CC       7S RNA and mediates binding of the 54 kDa subunit of the SRP (By
CC       similarity).
CC   -!- SUBUNIT: Archaeal signal recognition particle consists of a 7S RNA
CC       molecule of 300 nucleotides and two protein subunits: SRP54 and
CC       SRP19 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the SRP19 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE010688; AAM03745.1; ALT_INIT.
DR   HAMAP; MF_00305; -; 1.
DR   InterPro; IPR002778; SRP19.
DR   Pfam; PF01922; SRP19; 1.
DR   ProDom; PD006609; SRP19; 1.
KW   Signal recognition particle; RNA-binding; Ribonucleoprotein;
KW   Complete proteome.
SQ   SEQUENCE   101 AA;  11415 MW;  8DA2E31AAA9594C3 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 101;
  Best Local Similarity   100.0%;  Pred. No. 46;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         79 SVIAK 83


RESULT 4
SR19_METMA
ID   SR19_METMA     STANDARD;      PRT;   101 AA.
AC   Q8PWM7;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Signal recognition particle 19 kDa protein (SRP19).
GN   SRP19 OR MM1557.
```

```
OS   Methanosarcina mazei (Methanosarcina frisia).
OC   Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC   Methanosarcinaceae; Methanosarcina.
OX   NCBI_TaxID=2209;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Goe1 / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX   MEDLINE=22120827; PubMed=12125824;
RA   Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA   Martinez-Arias R., Henne A., Wiezer A., Baeumer S., Jacobi C.,
RA   Brueggemann H., Lienard T., Christmann A., Boemecke M., Steckel S.,
RA   Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA   Fritz H.-J., Gottschalk G.;
RT   "The genome of Methanosarcina mazei: evidence for lateral gene
RT   transfer between Bacteria and Archaea.";
RL   J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
CC   -!- FUNCTION: Signal-recognition-particle assembly, binds directly to
CC       7S RNA and mediates binding of the 54 kDa subunit of the SRP (By
CC       similarity).
CC   -!- SUBUNIT: Archaeal signal recognition particle consists of a 7S RNA
CC       molecule of 300 nucleotides and two protein subunits: SRP54 and
CC       SRP19 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the SRP19 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE013390; AAM31253.1; -.
DR   HAMAP; MF_00305; -; 1.
DR   InterPro; IPR002778; SRP19.
DR   Pfam; PF01922; SRP19; 1.
DR   ProDom; PD006609; SRP19; 1.
KW   Signal recognition particle; RNA-binding; Ribonucleoprotein;
KW   Complete proteome.
SQ   SEQUENCE   101 AA;  11377 MW;  3F9235C41CF68C74 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 101;
  Best Local Similarity   100.0%;  Pred. No. 46;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         79 SVIAK 83


RESULT 5
YA7B_HAEIN
ID   YA7B_HAEIN     STANDARD;      PRT;   134 AA.
AC   Q57425; P96338;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein HI1077.1.
GN   HI1077.1.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: WEAK, TO BACTERIAL PNUC PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32788; AAC22744.1; -.
DR   TIGR; HI1077.1; -.
DR   InterPro; IPR006419; NMN_trans_PnuC.
DR   Pfam; PF04973; NMN_transporter; 1.
DR   TIGRFAMs; TIGR01528; NMN_trans_PnuC; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     23     43       POTENTIAL.
FT   TRANSMEM     81    101       POTENTIAL.
FT   TRANSMEM    113    133       POTENTIAL.
SQ   SEQUENCE   134 AA;  14415 MW;  70C1620F88D0E6BF CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 134;
  Best Local Similarity   100.0%;  Pred. No. 60;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         75 SVIAK 79


RESULT 6
PUR3_STAAM
ID   PUR3_STAAM     STANDARD;      PRT;   188 AA.
AC   Q99V25;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Phosphoribosylglycinamide formyltransferase (EC 2.1.2.2) (GART) (GAR
DE   transformylase) (5'-phosphoribosylglycinamide transformylase).
GN   PURN OR SAV1072 OR SA0924.
OS   Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS   Staphylococcus aureus (strain N315).
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=158878, 158879;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Mu50 / ATCC 700699, and N315;
RX   MEDLINE=21311952; PubMed=11418146;
RA   Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA   Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA   Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA   Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA   Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA   Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA   Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT   "Whole genome sequencing of meticillin-resistant Staphylococcus
RT   aureus.";
RL   Lancet 357:1225-1240(2001).
CC   -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + N(1)-(5-phospho-D-
CC       ribosyl)glycinamide = tetrahydrofolate + N(2)-formyl-N(1)-(5-
CC       phospho-D-ribosyl)glycinamide.
CC   -!- PATHWAY: De novo purine biosynthesis; third step.
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- SIMILARITY: TO OTHER GART FROM BACTERIA AND EUKARYOTES.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AP003361; BAB57234.1; -.
DR   EMBL; AP003132; BAB42169.1; -.
DR   PIR; F89876; F89876.
DR   HSSP; P08179; 1GAR.
DR   SWISS-2DPAGE; Q99V25; STAAN.
DR   InterPro; IPR002376; formyl_transf.
DR   Pfam; PF00551; formyl_transf; 1.
DR   PROSITE; PS00373; GART; FALSE_NEG.
KW   Purine biosynthesis; Transferase; Complete proteome.
FT   ACT_SITE    146    146       BY SIMILARITY.
SQ   SEQUENCE   188 AA;  21166 MW;  F0364618F275FA30 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 188;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db        181 SVIAK 185


RESULT 7
PUR3_STAAW
ID   PUR3_STAAW     STANDARD;      PRT;   188 AA.
AC   Q8NX89;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Phosphoribosylglycinamide formyltransferase (EC 2.1.2.2) (GART) (GAR
DE   transformylase) (5'-phosphoribosylglycinamide transformylase).
GN   PURN OR MW0955.
OS   Staphylococcus aureus (strain MW2).
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=196620;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22040717; PubMed=12044378;
RA   Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA   Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA   Yamamoto K., Hiramatsu K.;
RT   "Genome and virulence determinants of high virulence community-
RT   acquired MRSA.";
RL   Lancet 359:1819-1827(2002).
CC   -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + N(1)-(5-phospho-D-
CC       ribosyl)glycinamide = tetrahydrofolate + N(2)-formyl-N(1)-(5-
CC       phospho-D-ribosyl)glycinamide.
CC   -!- PATHWAY: De novo purine biosynthesis; third step.
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- SIMILARITY: TO OTHER GART FROM BACTERIA AND EUKARYOTES.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AP004825; BAB94820.1; -.
DR   InterPro; IPR002376; formyl_transf.
DR   Pfam; PF00551; formyl_transf; 1.
DR   PROSITE; PS00373; GART; FALSE_NEG.
KW   Purine biosynthesis; Transferase; Complete proteome.
FT   ACT_SITE    146    146       BY SIMILARITY.
SQ   SEQUENCE   188 AA;  21153 MW;  D034134258D89AEE CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 188;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db        181 SVIAK 185


RESULT 8
RNH2_SYNPX
ID   RNH2_SYNPX     STANDARD;      PRT;   224 AA.
AC   Q7U4C6;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Ribonuclease HII (EC 3.1.26.4) (RNase HII).
GN   RNHB OR SYNW2144.
OS   Synechococcus sp. (strain WH8102).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=84588;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22825697; PubMed=12917641;
RA   Palenik B., Brahamsha B., Larimer F.W., Land M., Hauser L., Chain P.,
RA   Lamerdin J., Regala W., Allen E.E., McCarren J., Paulsen I.,
RA   Dufresne A., Partensky F., Webb E.A., Waterbury J.;
RT   "The genome of a motile marine Synechococcus.";
RL   Nature 424:1037-1042(2003).
CC   -!- FUNCTION: This enzyme is an endonuclease that degrades the RNA of
CC       RNA-DNA hybrids specifically (By similarity).
CC   -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC       phosphomonoester.
CC   -!- COFACTOR: Manganese (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: Belongs to the RNase HII family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; BX569694; CAE08659.1; ALT_INIT.
DR   HAMAP; MF_00052; -; 1.
DR   InterPro; IPR001352; RNase_HII/HIII.
DR   Pfam; PF01351; RNase_HII; 1.
KW   Hydrolase; Nuclease; Endonuclease; Manganese; Complete proteome.
FT   ACT_SITE     42     42       BY SIMILARITY.
FT   ACT_SITE    138    138       BY SIMILARITY.
FT   ACT_SITE    157    157       BY SIMILARITY.
SQ   SEQUENCE   224 AA;  24310 MW;  7F21360ABC4C54AB CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 224;
  Best Local Similarity   100.0%;  Pred. No. 96;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db        167 SVIAK 171


RESULT 9
VGLL_HSV2H
ID   VGLL_HSV2H     STANDARD;      PRT;   224 AA.
AC   P28278;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Glycoprotein L precursor.
```

```
GN   GL OR UL1.
OS   Herpes simplex virus (type 2 / strain HG52).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10315;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92113549; PubMed=1662697;
RA   McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;
RT   "Comparative sequence analysis of the long repeat regions and
RT   adjoining parts of the long unique regions in the genomes of herpes
RT   simplex viruses types 1 and 2.";
RL   J. Gen. Virol. 72:3057-3075(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Dolan A.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ASSOCIATED WITH GLYCOPROTEIN H (GH) TO FORM A COMPLEX
CC       IMPORTANT FOR INFECTION AND CELL FUSION (BY SIMILARITY).
CC   -!- SIMILARITY: Belongs to the herpesviruses glycoprotein L family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10470; BAA01264.1; -.
DR   EMBL; Z86099; CAB06761.1; -.
DR   PIR; JQ1494; WMBEHG.
DR   InterPro; IPR007923; Herpes_UL1.
DR   Pfam; PF05259; Herpes_UL1; 1.
KW   Glycoprotein; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    224       GLYCOPROTEIN L.
FT   CARBOHYD    170    170       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   224 AA;  25192 MW;  CD585849250D7C1F CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 224;
  Best Local Similarity   100.0%;  Pred. No. 96;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         30 SVIAK 34


RESULT 10
RS3_THEAC
ID   RS3_THEAC      STANDARD;      PRT;   225 AA.
AC   Q9HIR5;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   30S ribosomal protein S3P.
GN   RPS3P OR TA1265.
OS   Thermoplasma acidophilum.
OC   Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC   Thermoplasmataceae; Thermoplasma.
OX   NCBI_TaxID=2303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 1728;
RX   MEDLINE=20479972; PubMed=11029001;
RA   Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA   Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT   "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT   acidophilum.";
RL   Nature 407:508-513(2000).
CC   -!- FUNCTION: Binds the lower part of the 30S subunit head (By
CC       similarity).
CC   -!- SUBUNIT: Part of the 30S ribosomal subunit.
CC   -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.
CC   -!- SIMILARITY: Contains 1 KH type-2 domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL445067; CAC12389.1; -.
DR   HAMAP; MF_01309; -; 1.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR009019; KH_prok.
DR   InterPro; IPR004044; KH_TYPE_2.
DR   InterPro; IPR001351; Ribosomal_S3_C.
DR   InterPro; IPR005703; S3_euk_arch.
DR   Pfam; PF00013; KH; 1.
DR   Pfam; PF00189; Ribosomal_S3_C; 1.
DR   SMART; SM00322; KH; 1.
DR   TIGRFAMs; TIGR01008; rpsC_E_A; 1.
DR   PROSITE; PS50823; KH_TYPE_2; 1.
DR   PROSITE; PS00548; RIBOSOMAL_S3; FALSE_NEG.
KW   Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT   DOMAIN       16     85       KH TYPE-2.
SQ   SEQUENCE   225 AA;  24726 MW;  FE2B2B220091017F CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         92 SVIAK 96


RESULT 11
RS3_THEVO
ID   RS3_THEVO      STANDARD;      PRT;   225 AA.
AC   Q97BX1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   30S ribosomal protein S3P.
GN   RPS3P OR TV0334 OR TVG0336522.
OS   Thermoplasma volcanium.
OC   Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC   Thermoplasmataceae; Thermoplasma.
OX   NCBI_TaxID=50339;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GSS1 / DSM 4299 / JCM 9571;
RX   MEDLINE=20570466; PubMed=11121031;
RA   Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA   Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA   Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
RT   "Archaeal adaptation to higher temperatures revealed by genomic
RT   sequence of Thermoplasma volcanium.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
CC   -!- FUNCTION: Binds the lower part of the 30S subunit head (By
CC       similarity).
CC   -!- SUBUNIT: Part of the 30S ribosomal subunit.
CC   -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.
CC   -!- SIMILARITY: Contains 1 KH type-2 domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AP000992; BAB59476.1; -.
DR   HAMAP; MF_01309; -; 1.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR009019; KH_prok.
DR   InterPro; IPR004044; KH_TYPE_2.
DR   InterPro; IPR001351; Ribosomal_S3_C.
DR   InterPro; IPR005703; S3_euk_arch.
DR   Pfam; PF00013; KH; 1.
DR   Pfam; PF00189; Ribosomal_S3_C; 1.
DR   SMART; SM00322; KH; 1.
DR   TIGRFAMs; TIGR01008; rpsC_E_A; 1.
DR   PROSITE; PS50823; KH_TYPE_2; 1.
DR   PROSITE; PS00548; RIBOSOMAL_S3; FALSE_NEG.
KW   Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
FT   DOMAIN       16      85       KH TYPE-2.
SQ   SEQUENCE   225 AA;  24800 MW;  45BDBBFB26F52899 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         92 SVIAK 96


RESULT 12
RNH2_STRAW
ID   RNH2_STRAW     STANDARD;      PRT;   233 AA.
AC   Q82KF0;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Ribonuclease HII (EC 3.1.26.4) (RNase HII).
GN   RNHB OR RNH OR SAV2453.
OS   Streptomyces avermitilis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX   MEDLINE=22608306; PubMed=12692562;
RA   Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA   Sakaki Y., Hattori M., Omura S.;
RT   "Complete genome sequence and comparative analysis of the industrial
RT   microorganism Streptomyces avermitilis.";
RL   Nat. Biotechnol. 21:526-531(2003).
CC   -!- FUNCTION: This enzyme is an endonuclease that degrades the RNA of
CC       RNA-DNA hybrids specifically (By similarity).
CC   -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC       phosphomonoester.
CC   -!- COFACTOR: Manganese (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: Belongs to the RNase HII family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AP005030; BAC70164.1; -.
DR   HAMAP; MF_00052; -; 1.
DR   InterPro; IPR001352; RNase_HII/HIII.
DR   Pfam; PF01351; RNase_HII; 1.
KW   Hydrolase; Nuclease; Endonuclease; Manganese; Complete proteome.
FT   ACT_SITE     27     27       BY SIMILARITY.
FT   ACT_SITE    119    119       BY SIMILARITY.
FT   ACT_SITE    138    138       BY SIMILARITY.
SQ   SEQUENCE   233 AA;  25206 MW;  ADD09E963161FE19 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 233;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db        148 SVIAK 152


RESULT 13
FLGD_BUCAI
ID   FLGD_BUCAI     STANDARD;      PRT;   236 AA.
AC   P57421;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Basal-body rod modification protein flgD.
GN   FLGD OR BU339.
OS   Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS   symbiotic bacterium).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Buchnera.
OX   NCBI_TaxID=118099;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Tokyo 1998;
RX   MEDLINE=20445173; PubMed=10993077;
RA   Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT   "Genome sequence of the endocellular bacterial symbiont of aphids
RT   Buchnera sp. APS.";
RL   Nature 407:81-86(2000).
CC   -!- FUNCTION: REQUIRED FOR FLAGELLAR HOOK FORMATION. MAY ACT AS A
CC       SCAFFOLDING PROTEIN (BY SIMILARITY).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AP001119; BAB13044.1; -.
DR   InterPro; IPR005648; FlgD.
DR   Pfam; PF03963; FlgD; 1.
KW   Flagellum; Complete proteome.
SQ   SEQUENCE   236 AA;  26187 MW;  E15EAA2D3D84F293 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 236;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db        186 SVIAK 190


RESULT 14
VNST_PTPV
```

```
ID   VNST_PTPV      STANDARD;      PRT;   250 AA.
AC   P03516;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nonstructural protein NS-S.
OS   Punta toro phlebovirus.
OC   Viruses; ssRNA negative-strand viruses; Bunyaviridae; Phlebovirus.
OX   NCBI_TaxID=11587;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84276006; PubMed=6087547;
RA   Ihara T., Akashi H., Bishop D.H.L.;
RT   "Novel coding strategy (ambisense genomic RNA) revealed by sequence
RT   analyses of Punta toro phlebovirus S RNA.";
RL   Virology 136:293-306(1984).
CC   -!- MISCELLANEOUS: This protein may be a transcriptase component.
CC   -!- SIMILARITY: NS-S FROM PUNTA TORO, RIFT VALLEY FEVER, SANDFLY FEVER
CC       SICILIAN, TOSCANA, AND UUKUNIEMI VIRUSES ARE EVOLUTIONARY RELATED.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K02736; AAA47115.1; -.
DR   PIR; A04108; MNVUPT.
KW   Nonstructural protein; Transcription.
SQ   SEQUENCE   250 AA;  29097 MW;  2C8909A1EDAD90D7 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 250;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         55 SVIAK 59


RESULT 15
EXBB_XANCP
ID   EXBB_XANCP     STANDARD;      PRT;   253 AA.
AC   O34260;
DT   15-DEC-1998 (Rel. 37, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Biopolymer transport exbB protein.
GN   EXBB OR XCC0009.
OS   Xanthomonas campestris (pv. campestris).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=340;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B100;
RX   MEDLINE=98037510; PubMed=9371459;
RA   Wiggerich H.G., Klauke B., Koeplin R., Priefer U.B., Puehler A.;
RT   "Unusual structure of the tonB-exb DNA region of Xanthomonas
RT   campestris pv. campestris: tonB, exbB, and exbD1 are essential for
RT   ferric iron uptake, but exbD2 is not.";
RL   J. Bacteriol. 179:7103-7110(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33913 / NCPPB 528;
RX   MEDLINE=22022145; PubMed=12024217;
RA   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA   Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA   Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA   Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA   Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA   Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA   Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA   Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA   Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA   Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA   Setubal J.C., Kitajima J.P.;
RT   "Comparison of the genomes of two Xanthomonas pathogens with differing
RT   host specificities.";
RL   Nature 417:459-463(2002).
CC   -!- FUNCTION: Involved in the tonB-dependent energy-dependent
CC       transport of various receptor-bound substrates. Protects exbD from
CC       proteolytic degradation and functionally stabilizes tonB (By
CC       similarity).
CC   -!- SUBUNIT: The accessory proteins exbB and exbD seem to form a
CC       complex with tonB (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: Belongs to the exbB / tolQ family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z95386; CAB08609.1; -.
DR   EMBL; AE012094; AAM39328.1; -.
DR   InterPro; IPR002898; MotA_ExbB.
DR   Pfam; PF01618; MotA_ExbB; 1.
KW   Transport; Protein transport; Transmembrane; Inner membrane;
KW   Complete proteome.
FT   TRANSMEM     39     59       POTENTIAL.
FT   TRANSMEM    163    183       POTENTIAL.
FT   TRANSMEM    204    224       POTENTIAL.
FT   CONFLICT    114    114       A -> R (IN REF. 1).
SQ   SEQUENCE   253 AA;  26666 MW;  90138F91BC714508 CRC64;

  Query Match             100.0%;  Score 21;  DB 1;  Length 253;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db        230 SVIAK 234


Search completed: August 12, 2004, 06:20:04
Job time : 4.30738 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 0.461066 Seconds
(without alignments)
1043.144 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	100.0	86	2	AI2955	cell division topo
2	21	100.0	86	2	D98327	cell division topo
3	21	100.0	132	2	T32373	hypothetical prote
4	21	100.0	171	2	G90532	hypothetical prote
5	21	100.0	181	2	AF1931	hypothetical prote
6	21	100.0	188	2	F89876	phosphoribosylglyc
7	21	100.0	194	2	T15115	hypothetical prote
8	21	100.0	197	2	D87309	hypothetical prote
9	21	100.0	200	2	AG1137	probable sugar-pho
10	21	100.0	203	2	C85288	hypothetical prote
11	21	100.0	203	2	T05519	hypothetical prote
12	21	100.0	219	2	A97665	tetR family bacter
13	21	100.0	219	2	AD2889	transcription regu
14	21	100.0	224	1	WMBEHG	UL1 protein - huma
15	21	100.0	236	2	D84969	basal-body rod mod
16	21	100.0	243	2	H95909	probable membrane-
17	21	100.0	250	1	MNVUPT	nonstructural prot
18	21	100.0	263	1	S73489	probable S-adenosy
19	21	100.0	266	2	G81674	conserved hypothet
20	21	100.0	268	2	D64217	ribosomal protein
21	21	100.0	271	1	ELRT2	pancreatic elastas
22	21	100.0	279	2	T50125	probable 1-acylgly
23	21	100.0	282	1	JQ0347	cytochrome c1 - Rh
24	21	100.0	293	2	S65582	mating type protei
25	21	100.0	301	2	A95910	probable glycosylt
26	21	100.0	305	2	H82684	acetyltransferase
27	21	100.0	331	2	F97767	hypothetical prote
28	21	100.0	335	2	S48061	DNA-directed DNA p
29	21	100.0	335	2	A27112	DNA-directed DNA p
30	21	100.0	342	2	JC5791	aquaporin 9 - huma
31	21	100.0	348	2	S11198	transforming prote
32	21	100.0	355	1	A55473	early switch prote
33	21	100.0	357	1	G69290	probable hexosyltr
34	21	100.0	358	2	C84713	probable dioxygena
35	21	100.0	367	2	T01017	probable MYB famil
36	21	100.0	396	2	B96601	hypothetical prote
37	21	100.0	399	2	T49186	hypothetical prote
38	21	100.0	417	2	B55473	early switch prote
39	21	100.0	418	2	A96601	hypothetical prote
40	21	100.0	420	2	AF2464	hypothetical prote
41	21	100.0	426	2	C55473	early switch prote
42	21	100.0	430	2	S15308	hypothetical prote
43	21	100.0	470	2	G85877	hypothetical prote
44	21	100.0	470	2	F91033	hypothetical prote
45	21	100.0	498	2	T30092	hypothetical prote

ALIGNMENTS

RESULT 1
AI2955
cell division topological specificity factor minE [imported] - Agrobacterium tumefaciens
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AI2955
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClella
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AI2955
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-86 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44063.1; PID:g17741627; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: minE
A;Map position: linear chromosome

Query Match 100.0%; Score 21; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 SVIAK 5
              |||||
Db         45 SVIAK 49
```

RESULT 2
D98327
cell division topological specificity factor [imported] - Agrobacterium tumefaciens (str
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: D98327
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: D98327
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-86 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK90142.1; PID:g15160139; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_3134
A;Map position: linear chromosome

```
  Query Match             100.0%;  Score 21;  DB 2;  Length 86;
  Best Local Similarity   100.0%;  Pred. No. 55;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         45 SVIAK 49


RESULT 3
T32373
hypothetical protein C01B12.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jun-2000
C;Accession: T32373
R;Scheet, P.; Maggi, L.
submitted to the EMBL Data Library, September 1997
A;Description: The sequence of C. elegans cosmid C01B12.
A;Reference number: Z21156
A;Accession: T32373
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-132 <SCH>
A;Cross-references: EMBL:AF025458; PIDN:AAB70973.1; GSPDB:GN00020; CESP:C01B12.7
A;Experimental source: strain Bristol N2; clone C01B12
C;Genetics:
A;Gene: CESP:C01B12.7
A;Map position: 2
A;Introns: 23/3; 90/2
C;Superfamily: Caenorhabditis elegans hypothetical protein C01B12.7

  Query Match             100.0%;  Score 21;  DB 2;  Length 132;
  Best Local Similarity   100.0%;  Pred. No. 86;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         25 SVIAK 29


RESULT 4
G90532
hypothetical protein MYPU_1670 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: G90532
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: G90532
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-171 <KUR>
A;Cross-references: GB:AL445566; PID:g14089580; PIDN:CAC13340.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_1670
A;Genetic code: SGC3

  Query Match             100.0%;  Score 21;  DB 2;  Length 171;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db        117 SVIAK 121


RESULT 5
AF1931
hypothetical protein alr1001 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AF1931
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AF1931
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-181 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB72958.1; PID:g17130347; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr1001

  Query Match             100.0%;  Score 21;  DB 2;  Length 181;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         37 SVIAK 41


RESULT 6
F89876
phosphoribosylglycinamide formyltransferase [imported] - Staphylococcus aureus (strain N
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 21-Jun-2002
C;Accession: F89876
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: F89876
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-188 <KUR>
A;Cross-references: GB:BA000018; PID:g13700873; PIDN:BAB42169.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: purN

  Query Match             100.0%;  Score 21;  DB 2;  Length 188;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db        181 SVIAK 185


RESULT 7
T15115
hypothetical protein ZC132.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15115
R;Bradshaw, H.; Devlin, K.
submitted to the EMBL Data Library, July 1997
A;Description: The sequence of C. elegans cosmid ZC132.
A;Reference number: Z18294
A;Accession: T15115
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-194 <BRA>
A;Cross-references: EMBL:AF014939; NID:g2275620; PID:g2275628; PIDN:AAB63931.1; GSPDB:GN
A;Experimental source: strain Bristol N2; clone ZC132
```

C;Genetics:
A;Gene: CESP:ZC132.9
A;Map position: 5
A;Introns: 135/3; 153/3

Query Match 100.0%; Score 21; DB 2; Length 194;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
||||||
Db 56 SVIAK 60

RESULT 8
D87309
hypothetical protein CC0485 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002
C;Accession: D87309
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: D87309
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-197 <STO>
A;Cross-references: GB:AE005673; NID:g13421662; PIDN:AAK22472.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC0485
C;Superfamily: 50S ribosomal protein L25

Query Match 100.0%; Score 21; DB 2; Length 197;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
||||||
Db 73 SVIAK 77

RESULT 9
AG1137
probable sugar-phosphate isomerase homolog lmo0502 [imported] - Listeria monocytogenes (
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AG1137
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AG1137
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-200 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC98581.1; PID:g16409878; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo0502

Query Match 100.0%; Score 21; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
||||||
Db 130 SVIAK 134

RESULT 10
C85288
hypothetical protein AT4g24980 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C;Accession: C85288
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488; PMID:10617198
A;Accession: C85288
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-203 <STO>
A;Cross-references: GB:NC_001268; NID:g7269348; PIDN:CAB79407.1; GSPDB:GN00140
C;Genetics:
A;Gene: AT4g24980
A;Map position: 4

Query Match 100.0%; Score 21; DB 2; Length 203;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
||||||
Db 30 SVIAK 34

RESULT 11
T05519
hypothetical protein F13M23.120 - Arabidopsis thaliana (fragment)
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C;Accession: T05519
R;Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.X
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15419
A;Accession: T05519
A;Molecule type: DNA
A;Residues: 1-203 <BEV>
A;Cross-references: EMBL:AL035523
A;Experimental source: cultivar Columbia; BAC clone F13M23
C;Genetics:
A;Map position: 4
A;Note: intron positions not resolved
A;Note: F13M23.120

Query Match 100.0%; Score 21; DB 2; Length 203;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
||||||
Db 30 SVIAK 34

RESULT 12
A97665
tetR family bacterial regulatory protein (AF232237) [imported] - Agrobacterium tumefacier
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: A97665
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: A97665
A;Status: preliminary
A;Molecule type: DNA

```
A;Residues: 1-219 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK88274.1; PID:g15157738; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_4617
A;Map position: circular chromosome

  Query Match             100.0%;  Score 21;  DB 2;  Length 219;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db        146 SVIAK 150


RESULT 13
AD2889
transcription regulator, TetR family ameR [imported] - Agrobacterium tumefaciens (strain
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AD2889
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AD2889
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-219 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL43530.1; PID:g17741041; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: ameR
A;Map position: circular chromosome

  Query Match             100.0%;  Score 21;  DB 2;  Length 219;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db        146 SVIAK 150


RESULT 14
WMBEHG
UL1 protein - human herpesvirus 2 (strain HG52)
C;Species: human herpesvirus 2
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jun-2000
C;Accession: JQ1494
R;McGeoch, D.J.; Cunningham, C.; McIntyre, G.; Dolan, A.
J. Gen. Virol. 72, 3057-3075, 1991
A;Title: Comparative sequence analysis of the long repeat regions and adjoining parts of
A;Reference number: JQ1494; MUID:92113549; PMID:1662697
A;Accession: JQ1494
A;Molecule type: DNA
A;Residues: 1-224 <MCG>
A;Cross-references: GB:D10470; DDBJ:D01127; NID:g221791; PIDN:BAA01264.1; PID:g221792
C;Genetics:
A;Gene: UL1
C;Superfamily: varicella-zoster virus gene 60 protein

  Query Match             100.0%;  Score 21;  DB 1;  Length 224;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         30 SVIAK 34


RESULT 15
D84969
basal-body rod modification protein flgD [imported] - Buchnera sp. (strain APS)
C;Species: Buchnera sp.
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C;Accession: D84969
R;Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A;Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A
A;Reference number: A84930; MUID:20445173; PMID:10993077
A;Accession: D84969
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-236 <STO>
A;Cross-references: GB:AP000398; GSPDB:GN00144
A;Experimental source: strain APS
C;Genetics:
A;Gene: flgD; BU339

  Query Match             100.0%;  Score 21;  DB 2;  Length 236;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db        186 SVIAK 190


Search completed: August 12, 2004, 06:13:48
Job time : 3.46107 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2004, 06:12:47 ; Search time 0.522541 Seconds
(without alignments)
493.990 Million cell updates/sec

Title: US-09-890-463-1
Perfect score: 21
Sequence: 1 SVIAK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	100.0	50	3	US-09-156-316-6	Sequence 6, Appli
2	21	100.0	52	1	US-08-519-103-14	Sequence 14, Appl
3	21	100.0	52	3	US-09-018-635-14	Sequence 14, Appl
4	21	100.0	52	4	US-09-912-962-14	Sequence 14, Appl
5	21	100.0	81	4	US-09-134-001C-4236	Sequence 4236, Ap
6	21	100.0	102	4	US-09-370-838-196	Sequence 196, App
7	21	100.0	111	4	US-09-107-532A-5388	Sequence 5388, Ap
8	21	100.0	156	3	US-08-928-941D-4	Sequence 4, Appli
9	21	100.0	156	3	US-08-928-941D-36	Sequence 36, Appl
10	21	100.0	156	4	US-09-280-590A-4	Sequence 4, Appli
11	21	100.0	156	4	US-09-280-590A-46	Sequence 46, Appl
12	21	100.0	156	4	US-09-892-398-4	Sequence 4, Appli
13	21	100.0	156	4	US-09-892-398-46	Sequence 46, Appl
14	21	100.0	212	4	US-09-489-039A-12172	Sequence 12172, A
15	21	100.0	244	4	US-09-107-532A-5886	Sequence 5886, Ap
16	21	100.0	249	1	US-08-680-726A-88	Sequence 88, Appl
17	21	100.0	249	3	US-09-092-409-88	Sequence 88, Appl
18	21	100.0	255	4	US-09-328-352-6414	Sequence 6414, Ap
19	21	100.0	332	4	US-09-252-991A-24865	Sequence 24865, A
20	21	100.0	342	3	US-09-381-810A-1	Sequence 1, Appli
21	21	100.0	354	4	US-09-328-352-7825	Sequence 7825, Ap
22	21	100.0	401	4	US-09-465-558-70	Sequence 70, Appl
23	21	100.0	436	4	US-09-543-681A-6760	Sequence 6760, Ap
24	21	100.0	529	4	US-09-323-998E-23	Sequence 23, Appl
25	21	100.0	529	4	US-09-323-998E-47	Sequence 47, Appl
26	21	100.0	529	4	US-09-323-998E-50	Sequence 50, Appl
27	21	100.0	529	4	US-09-323-998E-51	Sequence 51, Appl
28	21	100.0	598	4	US-09-134-000C-4957	Sequence 4957, Ap
29	21	100.0	713	3	US-09-335-409-11	Sequence 11, Appl
30	21	100.0	713	4	US-09-568-102-11	Sequence 11, Appl
31	21	100.0	713	4	US-09-567-969-11	Sequence 11, Appl
32	21	100.0	713	4	US-09-568-480-11	Sequence 11, Appl
33	21	100.0	713	4	US-09-568-486-11	Sequence 11, Appl
34	21	100.0	713	4	US-09-568-472-11	Sequence 11, Appl
35	21	100.0	713	4	US-09-567-899-11	Sequence 11, Appl
36	21	100.0	1504	4	US-09-328-352-7046	Sequence 7046, Ap
37	20	95.2	32	3	US-08-433-522A-16	Sequence 16, Appl
38	20	95.2	32	3	US-09-135-166-16	Sequence 16, Appl
39	20	95.2	32	3	US-08-942-046-16	Sequence 16, Appl
40	20	95.2	48	3	US-09-107-858-24	Sequence 24, Appl
41	20	95.2	48	4	US-09-579-174-24	Sequence 24, Appl
42	20	95.2	59	1	US-08-485-455D-71	Sequence 71, Appl
43	20	95.2	59	2	US-08-482-130C-71	Sequence 71, Appl
44	20	95.2	59	2	US-08-484-211C-71	Sequence 71, Appl
45	20	95.2	59	3	US-08-906-769-71	Sequence 71, Appl

ALIGNMENTS

RESULT 1
US-09-156-316-6
; Sequence 6, Application US/09156316
; Patent No. 6183961
; GENERAL INFORMATION:
; APPLICANT: Bernstein, Harold S.
; APPLICANT: Coughlin, Shaun R.
; TITLE OF INVENTION: Methods and Compositions for Regulating Cell Cycle
; TITLE OF INVENTION: Progression
; FILE REFERENCE: UCSF-020/01US
; CURRENT APPLICATION NUMBER: US/09/156,316
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 60/060,688
; EARLIER FILING DATE: 1997-09-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-156-316-6

Query Match 100.0%; Score 21; DB 3; Length 50;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVIAK 5
|||||
Db 26 SVIAK 30

RESULT 2
US-08-519-103-14
; Sequence 14, Application US/08519103
; Patent No. 5733730
; GENERAL INFORMATION:
; APPLICANT: deLange, Titia
; TITLE OF INVENTION: TELOMERE REPEAT BINDING FACTOR AND
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USE THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KLAUBER & JACKSON
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/519,103
;       FILING DATE:  25-AUG-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Crane-Feury, Sharon E.
;       REGISTRATION NUMBER:  36,113
;       REFERENCE/DOCKET NUMBER:  600-1-142
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  201-487-5800
;       TELEFAX:  201-343-1684
;       TELEX:  133521
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  52 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-08-519-103-14

  Query Match             100.0%;  Score 21;  DB 1;  Length 52;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         27 SVIAK 31


RESULT 3
US-09-018-635-14
; Sequence 14, Application US/09018635
; Patent No. 6297356
;  GENERAL INFORMATION:
;    APPLICANT:  de Lange, Titia
;    APPLICANT:  Broccoli, Dominique
;    APPLICANT:  Smogorzewska, Agata
;    TITLE OF INVENTION:  TELOMERE REPEAT BINDING FACTOR AND
;    TITLE OF INVENTION:  DIAGNOSTIC AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KLAUBER & JACKSON
;      STREET:  411 Hackensack Avenue
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/018,635
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  David A. Jackson
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-142 CIP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  52 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-018-635-14

  Query Match             100.0%;  Score 21;  DB 3;  Length 52;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         27 SVIAK 31


RESULT 4
US-09-912-962-14
; Sequence 14, Application US/09912962
; Patent No. 6586577
;   GENERAL INFORMATION:
;       APPLICANT: de Lange, Titia
;                  Broccoli, Dominique
;                  Smogorzewska, Agata
;       TITLE OF INVENTION: TELOMERE REPEAT BINDING FACTOR AND
;                           DIAGNOSTIC AND THERAPEUTIC USE THEREOF
;       NUMBER OF SEQUENCES: 52
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: KLAUBER & JACKSON
;            STREET: 411 Hackensack Avenue
;            CITY: Hackensack
;            STATE: New Jersey
;            COUNTRY: USA
;            ZIP: 07601
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/912,962
;            FILING DATE: 25-Jul-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/018,635
;            FILING DATE: 04-FEB-1998
;       ATTORNEY/AGENT INFORMATION:
;            NAME: David A. Jackson
;            REGISTRATION NUMBER: 26,742
;            REFERENCE/DOCKET NUMBER: 600-1-142 CIP1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 201-487-5800
;            TELEFAX: 201-343-1684
;            TELEX: 133521
;   INFORMATION FOR SEQ ID NO: 14:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 52 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-912-962-14

  Query Match             100.0%;  Score 21;  DB 4;  Length 52;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         27 SVIAK 31


RESULT 5
US-09-134-001C-4236
; Sequence 4236, Application US/09134001C
; Patent No. 6380370
```

```
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4236
;   LENGTH: 81
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4236

  Query Match             100.0%;  Score 21;  DB 4;  Length 81;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         77 SVIAK 81


RESULT 6
US-09-370-838-196
; Sequence 196, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven G.
;  APPLICANT: Lodes, Michael J.
;  APPLICANT: Mohamath, Roadoh
;  APPLICANT: Secrist, Heather
;  TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
;  TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.475C1
;  CURRENT APPLICATION NUMBER: US/09/370,838
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/285,323
;  EARLIER FILING DATE: 1999-04-02
;  NUMBER OF SEQ ID NOS: 289
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 196
;   LENGTH: 102
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-370-838-196

  Query Match             100.0%;  Score 21;  DB 4;  Length 102;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         44 SVIAK 48


RESULT 7
US-09-107-532A-5388
; Sequence 5388, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,532A
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/085,598
;             FILING DATE: 14 May 1998
;             APPLICATION NUMBER: 60/051571
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-012
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 5388:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 111 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: YES
;        ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...111
;        SEQUENCE DESCRIPTION: SEQ ID NO: 5388:
US-09-107-532A-5388

  Query Match             100.0%;  Score 21;  DB 4;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         58 SVIAK 62


RESULT 8
US-08-928-941D-4
; Sequence 4, Application US/08928941D
; Patent No. 6180763
;  GENERAL INFORMATION:
;    APPLICANT:  Hirai, Hiroshi
;    APPLICANT:  Sherr, Charles
;    TITLE OF INVENTION:  CYCLIN-D BINDING FACTOR, AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Jackson, Esq.
;      STREET:  411 Hackensack Ave, Continental Plaza, 4th
;      STREET:  Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,941D
;      FILING DATE:
```

```
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  1340-1-002 N CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  156 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
US-08-928-941D-4

  Query Match             100.0%;  Score 21;  DB 3;  Length 156;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         79 SVIAK 83


RESULT 9
US-08-928-941D-36
; Sequence 36, Application US/08928941D
; Patent No. 6180763
;  GENERAL INFORMATION:
;    APPLICANT:  Hirai, Hiroshi
;    APPLICANT:  Sherr, Charles
;    TITLE OF INVENTION:  CYCLIN-D BINDING FACTOR, AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Jackson, Esq.
;      STREET:  411 Hackensack Ave, Continental Plaza, 4th
;      STREET:  Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,941D
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  1340-1-002 N CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  156 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  YES
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;      ORGANISM:  Gallus gallus
US-08-928-941D-36

  Query Match             100.0%;  Score 21;  DB 3;  Length 156;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         79 SVIAK 83


RESULT 10
US-09-280-590A-4
; Sequence 4, Application US/09280590A
; Patent No. 6303772
;  GENERAL INFORMATION:
;       APPLICANT: Hirai, Hiroshi
;                  Sherr, Charles
;                  Inoue, Kazushi
;                  Bodner, Sarah M.
;       TITLE OF INVENTION: CYCLIN-D BINDING FACTOR, AND USES
;                           THEREOF
;       NUMBER OF SEQUENCES: 46
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: David A. Jackson, Esq.
;            STREET: 411 Hackensack Ave, Continental Plaza, 4th
;                    Floor
;            CITY: Hackensack
;            STATE: New Jersey
;            COUNTRY: USA
;            ZIP: 07601
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/280,590A
;            FILING DATE: 29-Mar-1999
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Jackson Esq., David A.
;            REGISTRATION NUMBER: 26,742
;            REFERENCE/DOCKET NUMBER: 1340-1-002 N CP2
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 201-487-5800
;            TELEFAX: 201-343-1684
;  INFORMATION FOR SEQ ID NO: 4:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 156 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: <Unknown>
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       HYPOTHETICAL: NO
;       FRAGMENT TYPE: internal
;       ORIGINAL SOURCE:
;            ORGANISM: Mus musculus
;       SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-280-590A-4

  Query Match             100.0%;  Score 21;  DB 4;  Length 156;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         79 SVIAK 83


RESULT 11
```

```
US-09-280-590A-46
; Sequence 46, Application US/09280590A
; Patent No. 6303772
;   GENERAL INFORMATION:
;        APPLICANT: Hirai, Hiroshi
;                   Sherr, Charles
;                   Inoue, Kazushi
;                   Bodner, Sarah M.
;        TITLE OF INVENTION: CYCLIN-D BINDING FACTOR, AND USES
;                            THEREOF
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: David A. Jackson, Esq.
;             STREET: 411 Hackensack Ave, Continental Plaza, 4th
;                     Floor
;             CITY: Hackensack
;             STATE: New Jersey
;             COUNTRY: USA
;             ZIP: 07601
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/280,590A
;             FILING DATE: 29-Mar-1999
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Jackson Esq., David A.
;             REGISTRATION NUMBER: 26,742
;             REFERENCE/DOCKET NUMBER: 1340-1-002 N CP2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 201-487-5800
;             TELEFAX: 201-343-1684
;   INFORMATION FOR SEQ ID NO: 46:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 156 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: YES
;        FRAGMENT TYPE: <Unknown>
;        ORIGINAL SOURCE:
;             ORGANISM: Gallus gallus
;        SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-280-590A-46

  Query Match             100.0%;  Score 21;  DB 4;  Length 156;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         79 SVIAK 83


RESULT 12
US-09-892-398-4
; Sequence 4, Application US/09892398
; Patent No. 6673902
;   GENERAL INFORMATION:
;        APPLICANT: Hirai, Hiroshi
;                   Sherr, Charles
;                   Inoue, Kazushi
;                   Bodner, Sarah M.
;        TITLE OF INVENTION: CYCLIN-D BINDING FACTOR, AND USES
;                            THEREOF
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: David A. Jackson, Esq.
;             STREET: 411 Hackensack Ave, Continental Plaza, 4th
;                     Floor
;             CITY: Hackensack
;             STATE: New Jersey
;             COUNTRY: USA
;             ZIP: 07601
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/892,398
;             FILING DATE: 27-Jun-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/280,590
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Jackson Esq., David A.
;             REGISTRATION NUMBER: 26,742
;             REFERENCE/DOCKET NUMBER: 1340-1-002 N CP2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 201-487-5800
;             TELEFAX: 201-343-1684
;   INFORMATION FOR SEQ ID NO: 4:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 156 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: NO
;        FRAGMENT TYPE: internal
;        ORIGINAL SOURCE:
;             ORGANISM: Mus musculus
;        SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-892-398-4

  Query Match             100.0%;  Score 21;  DB 4;  Length 156;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         79 SVIAK 83


RESULT 13
US-09-892-398-46
; Sequence 46, Application US/09892398
; Patent No. 6673902
;   GENERAL INFORMATION:
;        APPLICANT: Hirai, Hiroshi
;                   Sherr, Charles
;                   Inoue, Kazushi
;                   Bodner, Sarah M.
;        TITLE OF INVENTION: CYCLIN-D BINDING FACTOR, AND USES
;                            THEREOF
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: David A. Jackson, Esq.
;             STREET: 411 Hackensack Ave, Continental Plaza, 4th
;                     Floor
;             CITY: Hackensack
;             STATE: New Jersey
;             COUNTRY: USA
;             ZIP: 07601
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
```

```
;              APPLICATION NUMBER: US/09/892,398
;              FILING DATE: 27-Jun-2001
;              CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 09/280,590
;              FILING DATE: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Jackson Esq., David A.
;              REGISTRATION NUMBER: 26,742
;              REFERENCE/DOCKET NUMBER: 1340-1-002 N CP2
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 201-487-5800
;              TELEFAX: 201-343-1684
;    INFORMATION FOR SEQ ID NO: 46:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 156 amino acids
;              TYPE: amino acid
;              STRANDEDNESS: <Unknown>
;              TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         HYPOTHETICAL: YES
;         FRAGMENT TYPE: <Unknown>
;         ORIGINAL SOURCE:
;              ORGANISM: Gallus gallus
;         SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-892-398-46

  Query Match             100.0%;  Score 21;  DB 4;  Length 156;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         79 SVIAK 83


RESULT 14
US-09-489-039A-12172
; Sequence 12172, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 12172
;   LENGTH: 212
;   TYPE: PRT
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12172

  Query Match             100.0%;  Score 21;  DB 4;  Length 212;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db         36 SVIAK 40


RESULT 15
US-09-107-532A-5886
; Sequence 5886, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,532A
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/085,598
;             FILING DATE: 14 May 1998
;             APPLICATION NUMBER: 60/051571
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-012
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 5886:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 244 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: YES
;        ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...244
;        SEQUENCE DESCRIPTION: SEQ ID NO: 5886:
US-09-107-532A-5886

  Query Match             100.0%;  Score 21;  DB 4;  Length 244;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SVIAK 5
              |||||
Db        186 SVIAK 190


Search completed: August 12, 2004, 06:21:04
Job time : 1.52254 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 13, 2004, 11:19:50 ; Search time 77 Seconds
(without alignments)
6061.217 Million cell updates/sec

Title: US-09-890-463-6
Perfect score: 841
Sequence: 1 tccgttatcgctaaacagat.....aaaagcggccgctcgaatta 841

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

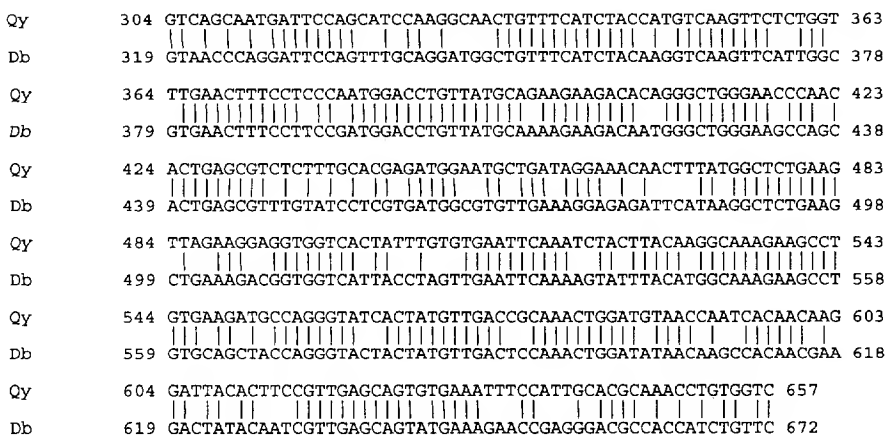
	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	383.6	45.6	678	4	US-09-459-956-6	Sequence 6, Appli
	2	362.6	43.1	699	4	US-09-459-956-5	Sequence 5, Appli
	3	274	32.6	801	4	US-09-459-956-7	Sequence 7, Appli
	4	198.4	23.6	690	4	US-09-459-956-2	Sequence 2, Appli
	5	154.2	18.3	696	4	US-09-459-956-3	Sequence 3, Appli
	6	151.8	18.0	696	4	US-09-459-956-4	Sequence 4, Appli
	7	131.6	15.6	720	4	US-09-839-650-1	Sequence 1, Appli
	8	126	15.0	1079	4	US-09-609-161B-15	Sequence 15, Appl
	9	126	15.0	1079	4	US-09-626-581D-64	Sequence 64, Appl
	10	126	15.0	1079	4	US-09-415-765B-64	Sequence 64, Appl
	11	126	15.0	1079	4	US-09-626-580C-64	Sequence 64, Appl
	12	126	15.0	1085	3	US-09-277-716-15	Sequence 15, Appl
	13	122.8	14.6	1021	4	US-09-839-650-2	Sequence 2, Appli
	14	122.6	14.6	1104	3	US-09-277-716-30	Sequence 30, Appl
	15	122.6	14.6	1104	4	US-09-609-161B-30	Sequence 30, Appl
	16	119.4	14.2	1279	3	US-09-277-716-31	Sequence 31, Appl
	17	119.4	14.2	1279	4	US-09-609-161B-31	Sequence 31, Appl
c	18	50.4	6.0	322	3	US-09-385-982-216	Sequence 216, App
c	19	50.4	6.0	322	3	US-09-385-982-362	Sequence 362, App
c	20	49.6	5.9	396	4	US-09-640-173-53	Sequence 53, Appl
c	21	49.6	5.9	396	4	US-09-713-550-53	Sequence 53, Appl
	22	48.4	5.8	6412	4	US-09-769-987-1	Sequence 1, Appli
	23	47.8	5.7	3275	4	US-09-370-838-151	Sequence 151, App
c	24	46.8	5.6	7218	1	US-08-232-463-14	Sequence 14, Appl
	25	46.2	5.5	144	1	US-08-702-344-26	Sequence 26, Appl
	26	46	5.5	1141	4	US-09-800-729-78	Sequence 78, Appl
	27	46	5.5	1927	4	US-09-336-536-66	Sequence 66, Appl
	28	45.8	5.4	2394	4	US-09-800-729-33	Sequence 33, Appl
c	29	45.6	5.4	396	4	US-09-640-173-10	Sequence 10, Appl
c	30	45.6	5.4	396	4	US-09-713-550-10	Sequence 10, Appl
	31	45.6	5.4	1737	1	US-08-202-056-4	Sequence 4, Appli
	32	45.6	5.4	1737	1	US-08-076-093A-3	Sequence 3, Appli
	33	45.6	5.4	1737	1	US-08-701-265-3	Sequence 3, Appli
	34	45.6	5.4	1737	2	US-08-284-586-3	Sequence 3, Appli
	35	45.6	5.4	1737	2	US-08-805-478-3	Sequence 3, Appli
	36	45.6	5.4	1737	2	US-08-802-627A-3	Sequence 3, Appli
	37	45.6	5.4	1737	2	US-08-801-238-3	Sequence 3, Appli
	38	45.6	5.4	1737	2	US-08-801-228-3	Sequence 3, Appli
	39	45.6	5.4	1737	3	US-09-104-296-3	Sequence 3, Appli
	40	45.6	5.4	1737	5	PCT-US94-06380-2	Sequence 2, Appli
	41	45.6	5.4	1738	2	US-08-379-482A-2	Sequence 2, Appli
	42	45.4	5.4	6409	4	US-09-967-908A-1	Sequence 1, Appli
	43	45.2	5.4	194	4	US-09-621-976-9596	Sequence 9596, Ap
	44	44.8	5.3	674	4	US-09-620-405B-465	Sequence 465, App
	45	44.8	5.3	674	4	US-09-433-826B-465	Sequence 465, App

ALIGNMENTS

RESULT 1
US-09-459-956-6
; Sequence 6, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp
US-09-459-956-6

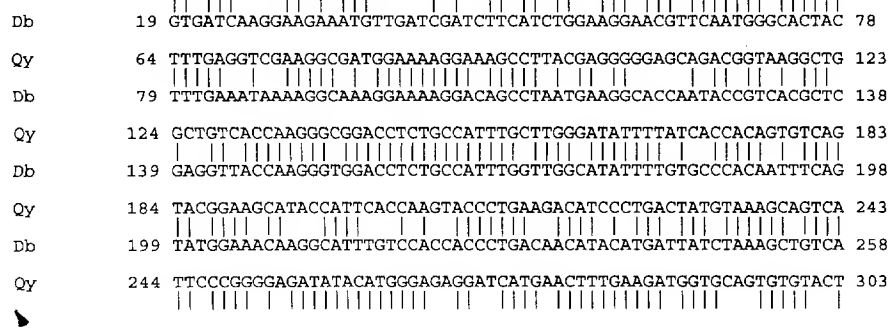
Query Match 45.6%; Score 383.6; DB 4; Length 678;
Best Local Similarity 74.2%; Pred. No. 1.4e-92;
Matches 485; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

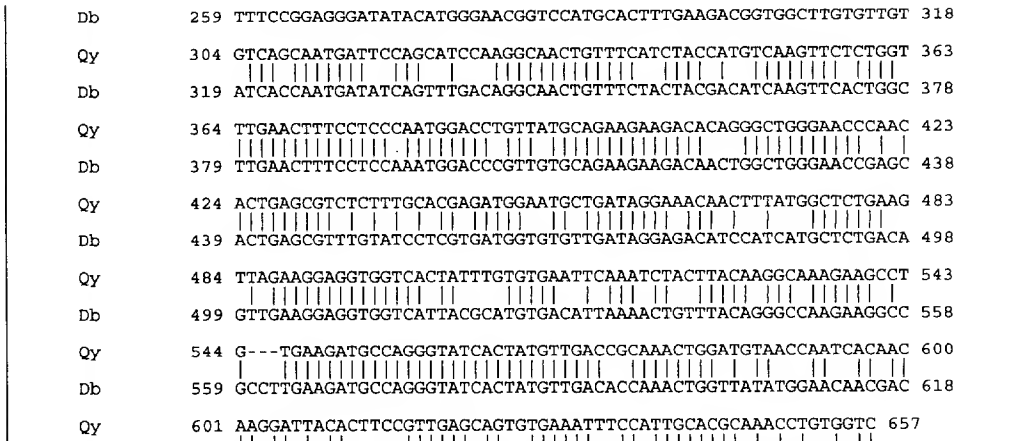
```
Qy          4 GTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACACTAC 63
              ||||||     |     ||||  |  || |||   |||  ||| ||||||||||| ||| |
Db         19 GTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGAACGGTCAATGGGCACGAG 78

Qy         64 TTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGGCTG 123
              |||||  | ||||||| ||| | || | ||| |||||  ||  |  | || |||| |||
Db         79 TTTGAAATAGAAGGCGAAGGAGAGGGGAGGCCATACGAAGGCCACAATACCGTAAAGCTT 138

Qy        124 GCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCAG 183
                || |||||||| |||||| ||||||||||||||||||||||| |||||||| | ||||
Db        139 AAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATTTTGTCACCACAATTTCAG 198

Qy        184 TACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCA 243
              || |||||||    ||     |||| ||||||  |||||  || ||||||  |||||  ||||
Db        199 TATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCAGACTATAAAAAGCTGTCA 258

Qy        244 TTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACT 303
              || ||  |   ||| || |||||| ||| ||||||||||||||| ||||   ||   ||||
Db        259 TTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAAGACGGTGGCGTCGTTACT 318
```



RESULT 2
US-09-459-956-5
; Sequence 5, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Discosoma striata
US-09-459-956-5

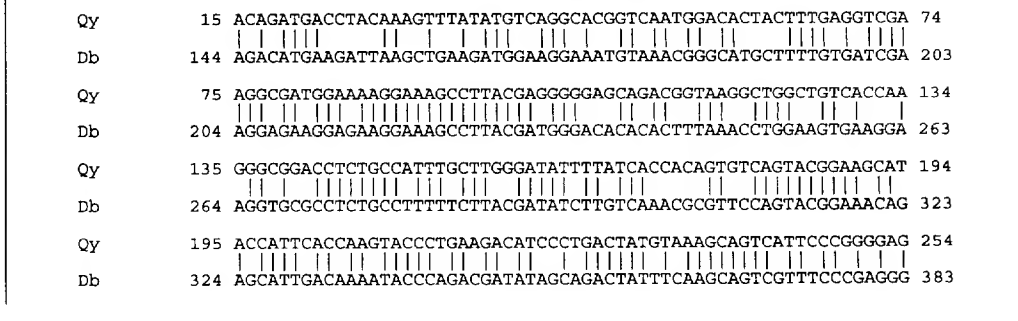
Query Match 43.1%; Score 362.6; DB 4; Length 699;
Best Local Similarity 73.1%; Pred. No. 5.3e-87;
Matches 480; Conservative 0; Mismatches 174; Indels 3; Gaps 1;

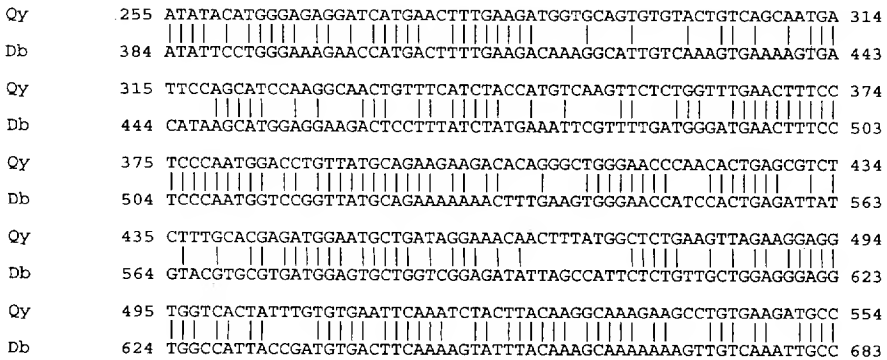




RESULT 3
US-09-459-956-7
; Sequence 7, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Clavularia sp
US-09-459-956-7

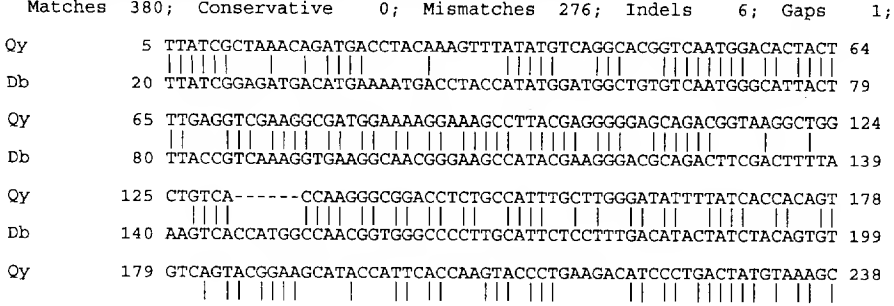
Query Match 32.6%; Score 274; DB 4; Length 801;
Best Local Similarity 64.2%; Pred. No. 1.9e-63;
Matches 412; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

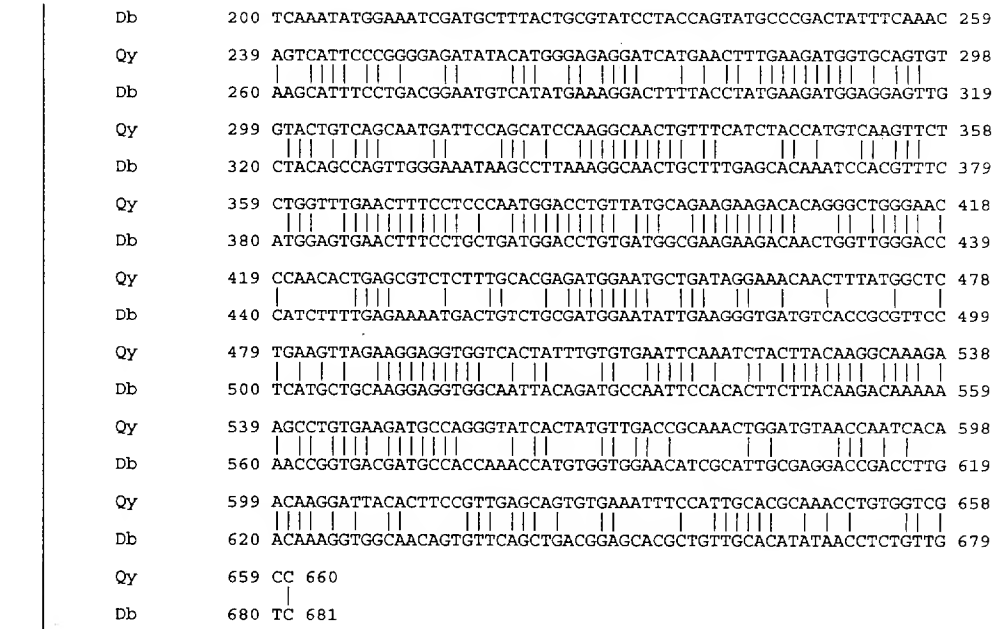




RESULT 4
US-09-459-956-2
; Sequence 2, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 690
; TYPE: DNA
; ORGANISM: Anemonia majano
US-09-459-956-2

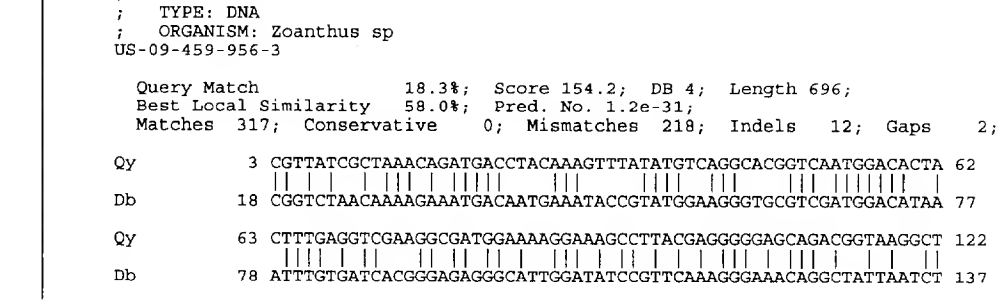
Query Match 23.6%; Score 198.4; DB 4; Length 690;
Best Local Similarity 57.4%; Pred. No. 2.1e-43;
Matches 380; Conservative 0; Mismatches 276; Indels 6; Gaps 1;





RESULT 5
US-09-459-956-3
; Sequence 3, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Zoanthus sp
US-09-459-956-3

Query Match 18.3%; Score 154.2; DB 4; Length 696;
Best Local Similarity 58.0%; Pred. No. 1.2e-31;
Matches 317; Conservative 0; Mismatches 218; Indels 12; Gaps 2;



```
Qy        123 GGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCA 182
Db        138 GTGTGTGGTCGAAGGTGGACCATTGCCATTTGCCGAAGACATATTGTCAGCTGCCTTTAA 197

Qy        183 GTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTC 242
Db        198 CTACGGAAACAGGGTTTTCACTGAATATCCTCAAGACATAGTTGACTATTTCAAGAACTC 257

Qy        243 ATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAG------T 296
Db        258 GTGTCCTGCTGGATATACATGGGACAGGTCTTTTCTCTTTGAGGATGGAGCAGTTTGCAT 317

Qy        297 GTGTACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTT 356
Db        318 ATGTAATGCAGATATAACAGTGAGTGTTGAAGAAAACTGCATGTATCATGAGTCCAAATT 377

Qy        357 CTCTGGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGA 416
Db        378 TTATGGAGTGAATTTTCCTGCTGATGGACCTGTGATGAAAAAGATGACAGATAACTGGGA 437

Qy        417 ACCCAACACTGAGCGTCTCTTTGCACGA------GATGGAATGCTGATAGGAAACAACTT 470
Db        438 GCCATCCTGCGAGAAGATCATACCAGTACCTAAGCAGGGGATATTGAAAGGGGATGTCTC 497

Qy        471 TATGGCTCTGAAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAA 530
Db        498 CATGTACCTCCTTCTGAAGGATGGTGGGCGTTTACGGTGCCAATTCGACACAGTTTACAA 557

Qy        531 GGCAAAG 537
Db        558 AGCAAAG 564
```

RESULT 6
US-09-459-956-4
; Sequence 4, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Zoanthus sp
US-09-459-956-4

Query Match 18.0%; Score 151.8; DB 4; Length 696;
Best Local Similarity 57.6%; Pred. No. 5e-31;
Matches 336; Conservative 0; Mismatches 232; Indels 15; Gaps 3;

```
Qy         14 AACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCG 73
Db         29 AAGAAATGACAATGAAATACCACATGGAAGGGTGCGTCAACGGACATAAATTTGTGATCA 88

Qy         74 AAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGGCTGGCTGTCACCA 133
Db         89 CGGGCGAAGGCATTGGATATCCGTTCAAAGGGAAACAGACTATTAATCTGTGTGTGATCG 148

Qy        134 AGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCAGTACGGAAGCA 193
```

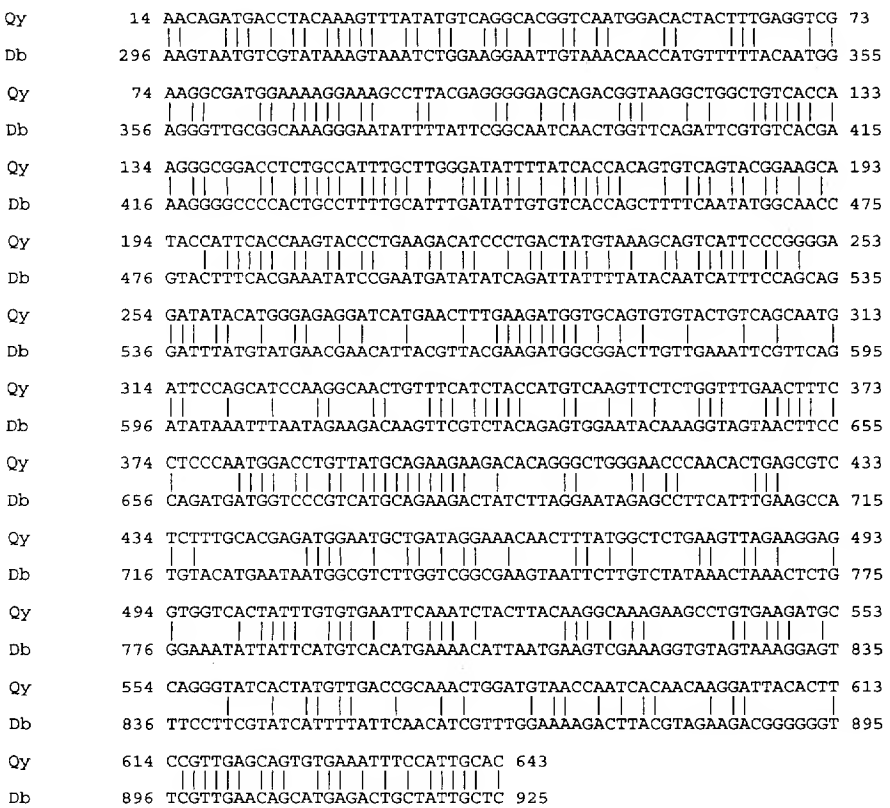
```
Db        149 AAGGGGGACCATTGCCATTTTCCGAAGACATATTGTCAGCTGGCTTTAAGTACGGAGACA 208

Qy        194 TACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCATTCCCGGGGA 253
Db        209 GGATTTTCACTGAATATCCTCAAGACATAGTAGACTATTTCAAGAACTCGTGTCCTGCTG 268

Qy        254 GATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCAATG 313
Db        269 GATATACATGGGGCAGGTCTTTTCTCTTTGAGGATGGAGCAGTCTGCATATGCAATGTAG 328

Qy        314 AT------TCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTGA 367
Db        329 ATATAACAGTGAGTGTCAAAGAAAACTGCATTTATCATAAGAGCATATTTAATGGAATGA 388

Qy        368 ACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAACACTG 427
Db        389 ATTTTCCTGCTGATGGACCTGTGATGAAAAAGATGACAACTAACTGGGAAGCATCCTGCG 448

Qy        428 AGCGTCTCTTTGCACGA------GATGGAATGCTGATAGGAAACAACTTTATGGCTCTGA 481
Db        449 AGAAGATCATGCCAGTACCTAAGCAGGGGATACTGAAAGGGGATGTCTCCATGTACCTCC 508

Qy        482 AGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAG---A 538
Db        509 TTCTGAAGGATGGTGGGCGTTACCGGTGCCAGTTCGACACAGTTTACAAAGCAAAGTCTG 568

Qy        539 AGCCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACT 581
Db        569 TGCCAAGTAAGATGCCGGAGTGGCACTTCATCCAGCATAAGCT 611
```

RESULT 7
US-09-839-650-1
; Sequence 1, Application US/09839650
; Patent No. 6645761
; GENERAL INFORMATION:
; APPLICANT: Stratagene
; TITLE OF INVENTION: Humanized Polynucleotide Sequence Encoding Renilla Mulleri Green
; Patent No. 6645761
; TITLE OF INVENTION: Fluorescent Protein
; FILE REFERENCE: 25436/1755
; CURRENT APPLICATION NUMBER: US/09/839,650
; CURRENT FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 720
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Humanized R. mulleri polynucleotide
; NAME/KEY: misc_feature
; LOCATION: (1)..(720)
; OTHER INFORMATION: Humanized DNA sequence
US-09-839-650-1

Query Match 15.6%; Score 131.6; DB 4; Length 720;
Best Local Similarity 50.6%; Pred. No. 1.2e-25;
Matches 317; Conservative 0; Mismatches 309; Indels 0; Gaps 0;

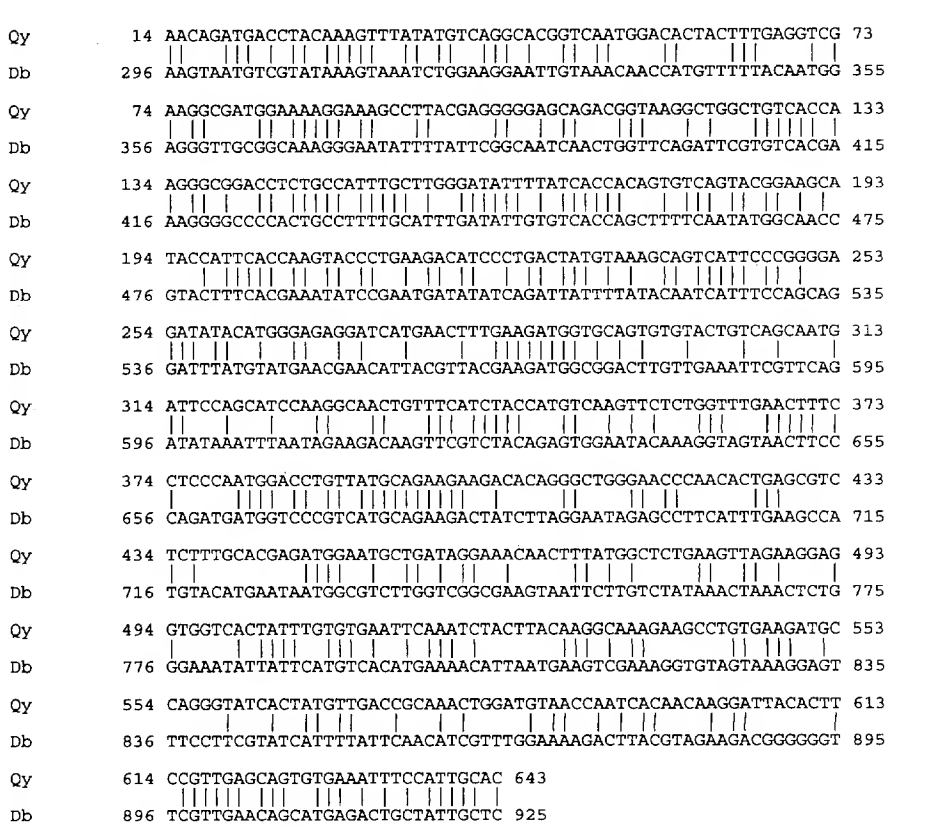
```
Qy         18 GATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCGAAGG 77
Db         45 GATGAGCTACAAGGTGAACCTGGAGGGCATCGTGAACAACCACGTGTTCACCATGGAGGG 104

Qy         78 CGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGGCTGGCTGTCACCAAGGG 137
Db        105 CTGCGGCAAGGGCAACATCCTGTTCGGCAACCAGCTGGTGCAGATCCGCGTGACCAAGGG 164

Qy        138 CGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCAGTACGGAAGCATACC 197
Db        165 CGCCCCCCTGCCCTTCGCCTTCGACATCGTGAGCCCCGCCTTCCAGTACGGCAACCGCAC 224

Qy        198 ATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCATTCCCGGGGAGATA 257
```

```
Db        225 CTTCACCAAGTACCCCAACGACATCAGCGACTACTTCATCCAGAGCTTCCCCGCCGGCTT 284

Qy        258 TACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCAATGATTC 317

Db        285 CATGTACGAGCGCACCCTGCGCTACGAGGACGGCGGCCTGGTGGAGATCCGCAGCGACAT 344

Qy        318 CAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTGAACTTTCCTCC 377

Db        345 CAACCTGATCGAGGACAAGTTCGTGTACCGCGTGGAGTACAAGGGCAGCAACTTCCCCGA 404

Qy        378 CAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAACACTGAGCGTCTCTT 437

Db        405 CGACGGCCCCGTGATGCAGAAGACCATCCTGGGCATCGAGCCCAGCTTCGAGGCCATGTA 464

Qy        438 TGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTGAAGTTAGAAGGAGGTGG 497

Db        465 CATGAACAACGGCGTGCTGGTGGGCGAGGTGATCCTGGTGTACAAGCTGAACAGCGGCAA 524

Qy        498 TCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAGCCTGTGAAGATGCCAGG 557

Db        525 GTACTACAGCTGCCACATGAAGACCCTGATGAAGAGCAAGGGCGTGGTGAAGGAGTTCCC 584

Qy        558 GTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAACAAGGATTACACTTCCGT 617

Db        585 CTCCTACCACTTCATCCAGCACCGCCTGGAGAAGACCTACGTGGAGGACGGCGGCTTCGT 644

Qy        618 TGAGCAGTGTGAAATTTCCATTGCAC 643

Db        645 GGAGCAGCACGAGACCGCCATCGCCC 670


RESULT 8
US-09-609-161B-15
; Sequence 15, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
;  APPLICANT: Bryan, Bruce
;  APPLICANT:  Szent-Gyorgyi, Christopher
;  APPLICANT:  PROLUME, LTD.
;  TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
;  TITLE OF INVENTION:  AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
;  TITLE OF INVENTION:  SCREENING AND NOVELTY ITEMS
;  FILE REFERENCE: 24729-121B
;  CURRENT APPLICATION NUMBER: US/09/609,161B
;  CURRENT FILING DATE:  2000-06-30
;  PRIOR APPLICATION NUMBER: 09/277,716
;  PRIOR FILING DATE: 1999-03-26
;  PRIOR APPLICATION NUMBER: 60/102,939
;  PRIOR FILING DATE: 1998-10-01
;  PRIOR APPLICATION NUMBER: 60/089,367
;  PRIOR FILING DATE: 1998-06-15
;  PRIOR APPLICATION NUMBER: 60/079,624
;  PRIOR FILING DATE: 1998-03-27
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
;   LENGTH: 1079
;   TYPE: DNA
;   ORGANISM: Renilla mulleri
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (259)..(975)
;   OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-609-161B-15

  Query Match           15.0%;  Score 126;  DB 4;  Length 1079;
  Best Local Similarity   50.0%;  Pred. No. 4.2e-24;
  Matches  315;  Conservative    0;  Mismatches  315;  Indels    0;  Gaps    0;

Qy         14 AACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCG 73

Db        296 AAGTAATGTCGTATAAAGTAAATCTGGAAGGAATTGTAAACAACCATGTTTTTACAATGG 355

Qy         74 AAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGGCTGGCTGTCACCA 133

Db        356 AGGGTTGCGGCAAAGGGAATATTTTATTCGGCAATCAACTGGTTCAGATTCGTGTCACGA 415

Qy        134 AGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCAGTACGGAAGCA 193

Db        416 AAGGGGCCCCACTGCCTTTTGCATTTGATATTGTGTCACCAGCTTTTCAATATGGCAACC 475

Qy        194 TACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCATTCCCGGGGA 253

Db        476 GTACTTTCACGAAATATCCGAATGATATATCAGATTATTTTATACAATCATTTCCAGCAG 535

Qy        254 GATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCAATG 313

Db        536 GATTTATGTATGAACGAACATTACGTTACGAAGATGGCGGACTTGTTGAAATTCGTTCAG 595

Qy        314 ATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTGAACTTTC 373

Db        596 ATATAAATTTAATAGAAGACAAGTTCGTCTACAGAGTGGAATACAAAGGTAGTAACTTCC 655

Qy        374 CTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAACACTGAGCGTC 433

Db        656 CAGATGATGGTCCCGTCATGCAGAAGACTATCTTAGGAATAGAGCCTTCATTTGAAGCCA 715

Qy        434 TCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTGAAGTTAGAAGGAG 493

Db        716 TGTACATGAATAATGGCGTCTTGGTCGGCGAAGTAATTCTTGTCTATAAACTAAACTCTG 775

Qy        494 GTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAGCCTGTGAAGATGC 553

Db        776 GGAAATATTATTCATGTCACATGAAAACATTAATGAAGTCGAAAGGTGTAGTAAAGGAGT 835

Qy        554 CAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAACAAGGATTACACTT 613

Db        836 TTCCTTCGTATCATTTTATTCAACATCGTTTGGAAAAGACTTACGTAGAAGACGGGGGGT 895

Qy        614 CCGTTGAGCAGTGTGAAATTTCCATTGCAC 643

Db        896 TCGTTGAACAGCATGAGACTGCTATTGCTC 925


RESULT 9
US-09-626-581D-64
; Sequence 64, Application US/09626581D
; Patent No. 6548249
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David
;  TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
;  TITLE OF INVENTION:  Libraries
;  FILE REFERENCE: A-66900-3/RMS
;  CURRENT APPLICATION NUMBER: US/09/626,581D
;  CURRENT FILING DATE:  2000-07-27
;  PRIOR APPLICATION NUMBER: 09/169,015
;  PRIOR FILING DATE: 1998-10-08
;  PRIOR APPLICATION NUMBER: 09/415,765
;  PRIOR FILING DATE: 1999-10-08
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
;   LENGTH: 1079
;   TYPE: DNA
;   ORGANISM: Renilla muelleri
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (259)..(975)
;   OTHER INFORMATION:
US-09-626-581D-64

  Query Match           15.0%;  Score 126;  DB 4;  Length 1079;
  Best Local Similarity   50.0%;  Pred. No. 4.2e-24;
  Matches  315;  Conservative    0;  Mismatches  315;  Indels    0;  Gaps    0;
```



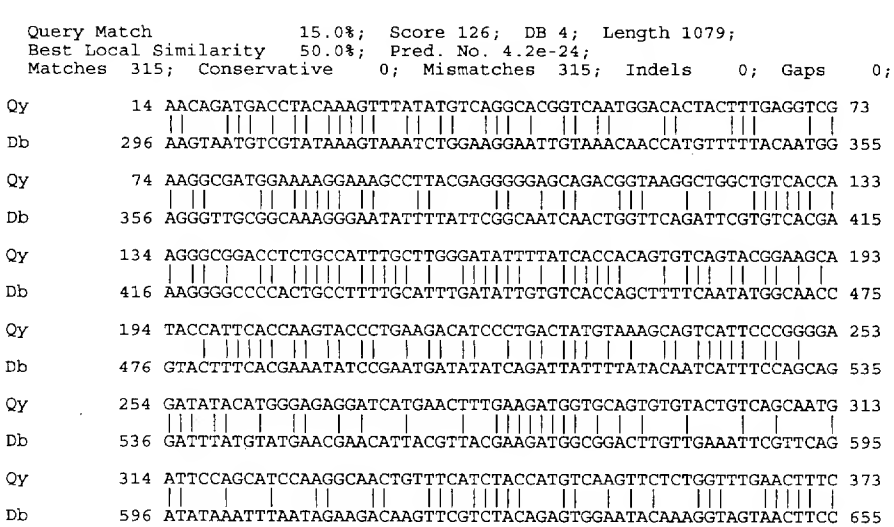
RESULT 10
US-09-415-765B-64
; Sequence 64, Application US/09415765B
; Patent No. 6548632
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A66900-1/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/415,765B
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Renilla muelleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; OTHER INFORMATION:
US-09-415-765B-64

Query Match 15.0%; Score 126; DB 4; Length 1079;
Best Local Similarity 50.0%; Pred. No. 4.2e-24;
Matches 315; Conservative 0; Mismatches 315; Indels 0; Gaps 0;



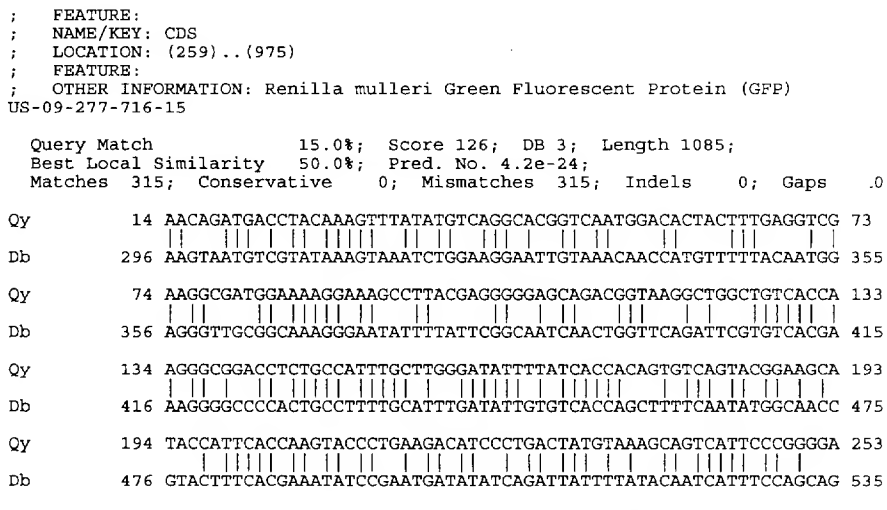
RESULT 11
US-09-626-580C-64
; Sequence 64, Application US/09626580C
; Patent No. 6562617
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Bogenberger, Jakob M.
; APPLICANT: Peele, Beau R.
; TITLE OF INVENTION: FUSIONS OF SCAFFOLD PROTEINS WITH RANDOM PEPTIDE LIBRARIES
; FILE REFERENCE: A-66900-2/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/626,580C
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 09/415,765
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Renilla muelleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; OTHER INFORMATION:
US-09-626-580C-64

Query Match 15.0%; Score 126; DB 4; Length 1079;
Best Local Similarity 50.0%; Pred. No. 4.2e-24;
Matches 315; Conservative 0; Mismatches 315; Indels 0; Gaps 0;



RESULT 12
US-09-277-716-15
; Sequence 15, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE L
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Renilla mulleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; FEATURE:
; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-277-716-15

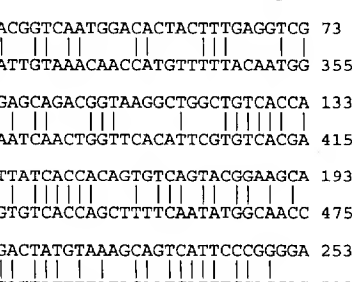
Query Match 15.0%; Score 126; DB 3; Length 1085;
Best Local Similarity 50.0%; Pred. No. 4.2e-24;
Matches 315; Conservative 0; Mismatches 315; Indels 0; Gaps 0;



RESULT 13
US-09-839-650-2
; Sequence 2, Application US/09839650
; Patent No. 6645761
; GENERAL INFORMATION:
; APPLICANT: Stratagene
; TITLE OF INVENTION: Humanized Polynucleotide Sequence Encoding Renilla Mulleri Green
; Patent No. 6645761
; TITLE OF INVENTION: Fluorescent Protein
; FILE REFERENCE: 25436/1755
; CURRENT APPLICATION NUMBER: US/09/839,650
; CURRENT FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1021
; TYPE: DNA

; ORGANISM: Renilla muelleri
; FEATURE:
; NAME/KEY: exon
; LOCATION: (259)..(976)
US-09-839-650-2

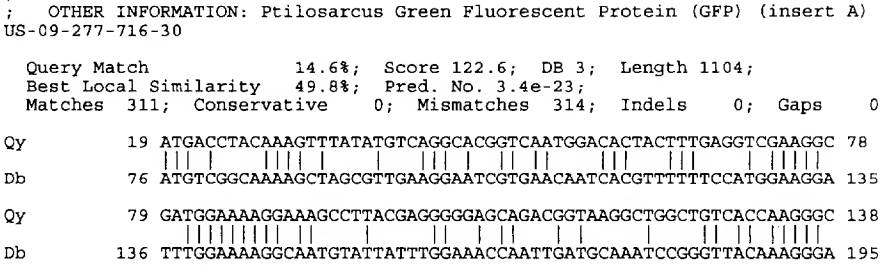
Query Match 14.6%; Score 122.8; DB 4; Length 1021;
Best Local Similarity 49.7%; Pred. No. 2.9e-23;
Matches 313; Conservative 0; Mismatches 317; Indels 0; Gaps 0;



RESULT 14
US-09-277-716-30
; Sequence 30, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34)..(747)
; FEATURE:
; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert A)
US-09-277-716-30

Query Match 14.6%; Score 122.6; DB 3; Length 1104;
Best Local Similarity 49.8%; Pred. No. 3.4e-23;
Matches 311; Conservative 0; Mismatches 314; Indels 0; Gaps 0;



RESULT 15
US-09-609-161B-30
; Sequence 30, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIGH
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B

```
;  CURRENT APPLICATION NUMBER: US/09/609,161B
;  CURRENT FILING DATE:  2000-06-30
;  PRIOR APPLICATION NUMBER: 09/277,716
;  PRIOR FILING DATE: 1999-03-26
;  PRIOR APPLICATION NUMBER: 60/102,939
;  PRIOR FILING DATE: 1998-10-01
;  PRIOR APPLICATION NUMBER: 60/089,367
;  PRIOR FILING DATE: 1998-06-15
;  PRIOR APPLICATION NUMBER: 60/079,624
;  PRIOR FILING DATE: 1998-03-27
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
;   LENGTH: 1104
;   TYPE: DNA
;   ORGANISM: Ptilosarcus gurneyi
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (34)..(747)
;   OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert A)
US-09-609-161B-30

  Query Match             14.6%;  Score 122.6;  DB 4;  Length 1104;
  Best Local Similarity   49.8%;  Pred. No. 3.4e-23;
  Matches  311;  Conservative    0;  Mismatches  314;  Indels    0;  Gaps    0;

Qy         19 ATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCGAAGGC 78
              ||| |   |||| |    |   ||| |  || ||   |||   |||    | |||||
Db         76 ATGTCGGCAAAAGCTAGCGTTGAAGGAATCGTGAACAATCACGTTTTTTCCATGGAAGGA 135

Qy         79 GATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGGCTGGCTGTCACCAAGGGC 138
              |||||||||  ||    |      ||  | ||   | |      |    || || |||||
Db        136 TTTGGAAAAGGCAATGTATTATTTGGAAACCAATTGATGCAAATCCGGGTTACAAAGGGA 195

Qy        139 GGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCAGTACGGAAGCATACCA 198
              ||  ||  ||||||| ||||  |||||| | ||   |   |  || ||||| |      |
Db        196 GGTCCGTTGCCATTCGCTTTCGATATTGTTTCCATAGCTTTCCAATACGGGAATCGCACT 255

Qy        199 TTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCATTCCCGGGGAGATAT 258
              ||||| || ||||| || |||||  | ||||| |    || ||||||||||   ||| |
Db        256 TTCACGAAATACCCAGACGACATTGCGGACTACTTTGTTCAATCATTCCCGGCTGGATTT 315

Qy        259 ACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCAATGATTCC 318
              |  || || |   |   |||||||||||| ||  |   |     |  |    |||
Db        316 TTCTACGAAAGAAATCTACGCTTTGAAGATGGCGCCATTGTTGACATTCGTTCAGATATA 375

Qy        319 AGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTGAACTTTCCTCCC 378
              ||  |  |||   |   |||  |||| | ||  |||     ||       || |||
Db        376 AGTTTAGAAGATGATAAGTTCCACTACAAAGTGGAGTATAGAGGCAACGGTTTCCCTAGT 435

Qy        379 AATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAACACTGAGCGTCTCTTT 438
              || ||||| || ||||| ||    |  |  |||  ||| ||      ||||    |||
Db        436 AACGGACCCGTGATGCAAAAAGCCATCCTCGGCATGGAGCCATCGTTTGAGGTGGTCTAC 495

Qy        439 GCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTGAAGTTAGAAGGAGGTGGT 498
                 ||  | ||| | ||  |       | | | |   ||  | ||   |||
Db        496 ATGAACAGCGGCGTTCTGGTGGGCGAAGTAGATCTCGTTTACAAACTCGAGTCAGGGAAC 555

Qy        499 CACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAGCCTGTGAAGATGCCAGGG 558
              | || | |||  |  | ||| |   |||||   | ||       |||||         |
Db        556 TATTACTCGTGCCACATGAAAACGTTTTACAGATCCAAAGGTGGAGTGAAAGAATTCCCG 615

Qy        559 TATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAACAAGGATTACACTTCCGTT 618
              |  |  |  ||  ||   | |     |  |  |   ||    | ||    |  | |||
Db        616 GAATATCACTTTATCCATCATCGTCTGGAGAAAACCTACGTGGAAGAAGGAAGCTTCGTG 675

Qy        619 GAGCAGTGTGAAATTTCCATTGCAC 643
              || ||    || |   |||||||||
Db        676 GAACAACACGAGACGGCCATTGCAC 700


Search completed: August 13, 2004, 20:48:16
Job time : 78 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 13, 2004, 11:19:50 ; Search time 77 Seconds
(without alignments)
6061.217 Million cell updates/sec

Title: US-09-890-463-5
Perfect score: 841
Sequence: 1 tccgttatcgctaaacagat.....aaaagcggccgctcgaatta 841

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

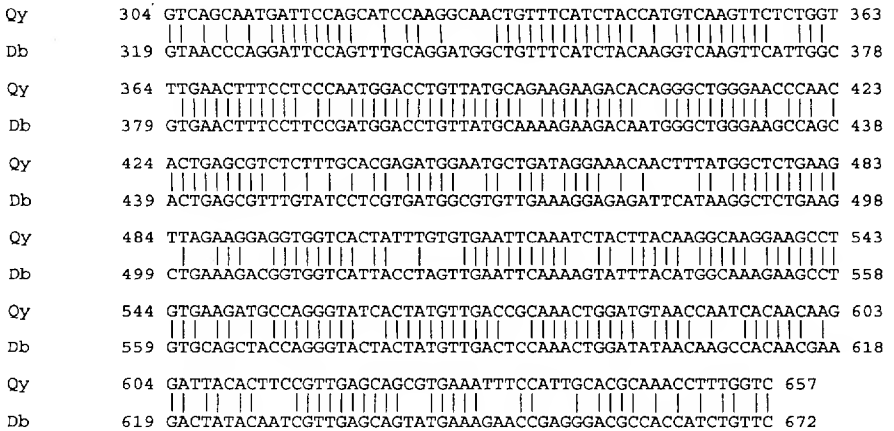
	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	380.4	45.2	678	4	US-09-459-956-6	Sequence 6, Appli
	2	362.6	43.1	699	4	US-09-459-956-5	Sequence 5, Appli
	3	272.4	32.4	801	4	US-09-459-956-7	Sequence 7, Appli
	4	198.4	23.6	690	4	US-09-459-956-2	Sequence 2, Appli
	5	152.6	18.1	696	4	US-09-459-956-3	Sequence 3, Appli
	6	150.2	17.9	696	4	US-09-459-956-4	Sequence 4, Appli
	7	131.6	15.6	720	4	US-09-839-650-1	Sequence 1, Appli
	8	126	15.0	1079	4	US-09-609-161B-15	Sequence 15, Appl
	9	126	15.0	1079	4	US-09-626-581D-64	Sequence 64, Appl
	10	126	15.0	1079	4	US-09-415-765B-64	Sequence 64, Appl
	11	126	15.0	1079	4	US-09-626-580C-64	Sequence 64, Appl
	12	126	15.0	1085	3	US-09-277-716-15	Sequence 15, Appl
	13	122.8	14.6	1021	4	US-09-839-650-2	Sequence 2, Appli
	14	122.6	14.6	1104	3	US-09-277-716-30	Sequence 30, Appl
	15	122.6	14.6	1104	4	US-09-609-161B-30	Sequence 30, Appl
	16	119.4	14.2	1279	3	US-09-277-716-31	Sequence 31, Appl
	17	119.4	14.2	1279	4	US-09-609-161B-31	Sequence 31, Appl
C	18	50.8	6.0	396	4	US-09-640-173-53	Sequence 53, Appl
C	19	50.8	6.0	396	4	US-09-713-550-53	Sequence 53, Appl
C	20	50.4	6.0	322	3	US-09-385-982-216	Sequence 216, App
C	21	50.4	6.0	322	3	US-09-385-982-362	Sequence 362, App
	22	49.4	5.9	3275	4	US-09-370-838-151	Sequence 151, App
	23	48.4	5.8	6412	4	US-09-769-987-1	Sequence 1, Appli
C	24	48	5.7	7218	1	US-08-232-463-14	Sequence 14, Appl
	25	47.8	5.7	2030	3	US-08-706-216-3	Sequence 3, Appli
	26	47.8	5.7	2030	4	US-09-650-284B-3	Sequence 3, Appli
	27	46	5.5	1712	4	US-09-148-545-106	Sequence 106, App
	28	46	5.5	1822	4	US-09-148-545-105	Sequence 105, App
	29	45.6	5.4	1737	1	US-08-202-056-4	Sequence 4, Appli
	30	45.6	5.4	1737	1	US-08-076-093A-3	Sequence 3, Appli
	31	45.6	5.4	1737	1	US-08-701-265-3	Sequence 3, Appli
	32	45.6	5.4	1737	2	US-08-284-586-3	Sequence 3, Appli
	33	45.6	5.4	1737	2	US-08-805-478-3	Sequence 3, Appli
	34	45.6	5.4	1737	2	US-08-802-627A-3	Sequence 3, Appli
	35	45.6	5.4	1737	2	US-08-801-238-3	Sequence 3, Appli
	36	45.6	5.4	1737	2	US-08-801-228-3	Sequence 3, Appli
	37	45.6	5.4	1737	3	US-09-104-296-3	Sequence 3, Appli
	38	45.6	5.4	1737	5	PCT-US94-06380-2	Sequence 2, Appli
C	39	45.2	5.4	47	2	US-08-778-494B-114	Sequence 114, App
	40	45	5.4	3828	4	US-09-221-013A-7	Sequence 7, Appli
	41	44.8	5.3	239	4	US-09-621-976-16632	Sequence 16632, A
	42	44.6	5.3	144	1	US-08-702-344-26	Sequence 26, Appl
	43	44.4	5.3	1141	4	US-09-800-729-78	Sequence 78, Appl
	44	44.4	5.3	1193	4	US-09-372-422A-23	Sequence 23, Appl
	45	44.4	5.3	1927	4	US-09-336-536-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1
US-09-459-956-6
; Sequence 6, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp
US-09-459-956-6

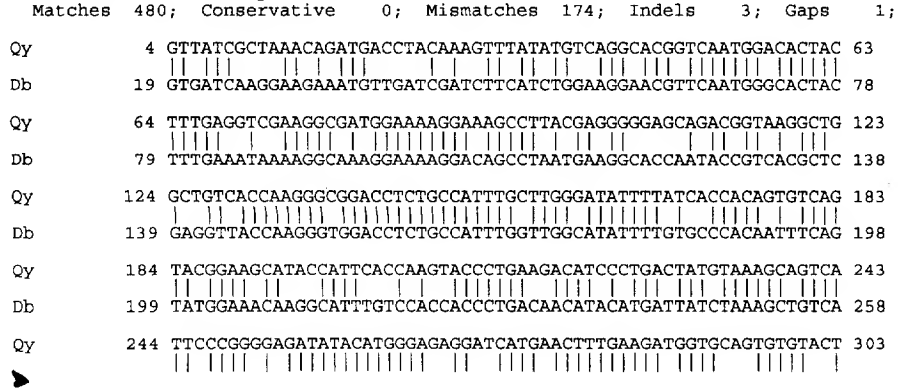
Query Match 45.2%; Score 380.4; DB 4; Length 678;
Best Local Similarity 73.9%; Pred. No. 4.8e-97;
Matches 483; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

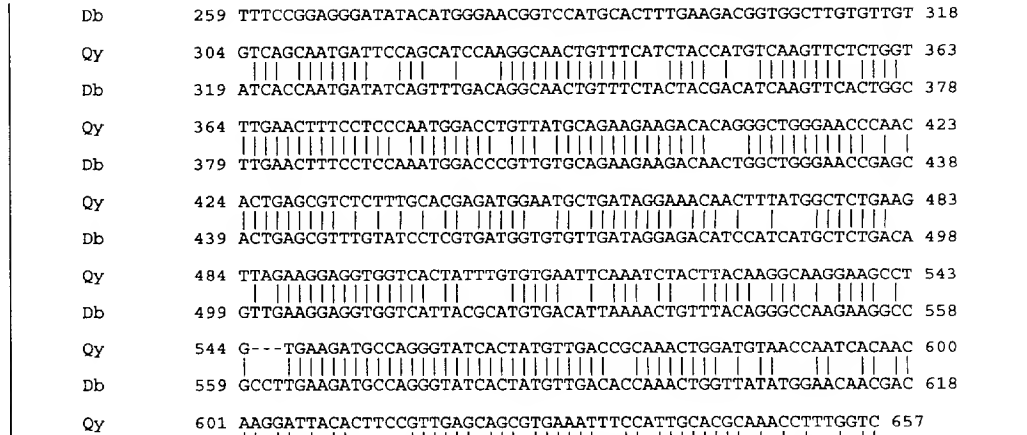
```
Qy          4 GTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACACTAC 63
              ||||||     |     |||| |  || |||  |||  ||| ||||||||||| ||| |
Db         19 GTTATCAAGGAGTTCATGAGGTTTAAGGTTCGCATGGAAGGAACGGTCAATGGGCACGAG 78

Qy         64 TTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGGCTG 123
              |||||  | ||||||| ||| | || | ||| |||||  ||  |  | || |||| |||
Db         79 TTTGAAATAGAAGGCGAAGGAGAGGGGAGGCCATACGAAGGCCACAATACCGTAAAGCTT 138

Qy        124 GCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCAG 183
                || |||||||| |||||| |||||||||||||||||||||| |||||||| | ||||
Db        139 AAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATTTTGTCACCACAATTTCAG 198

Qy        184 TACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCA 243
              || |||||||    ||     |||| |||||  ||||| || ||||||  ||||| ||||
Db        199 TATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCAGACTATAAAAAGCTGTCA 258

Qy        244 TTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACT 303
              || ||  |  ||| || |||||| ||| ||||||||||||||| ||||   ||   ||||
Db        259 TTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAAGACGGTGGCGTCGTTACT 318
```



RESULT 2
US-09-459-956-5
; Sequence 5, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Discosoma striata
US-09-459-956-5

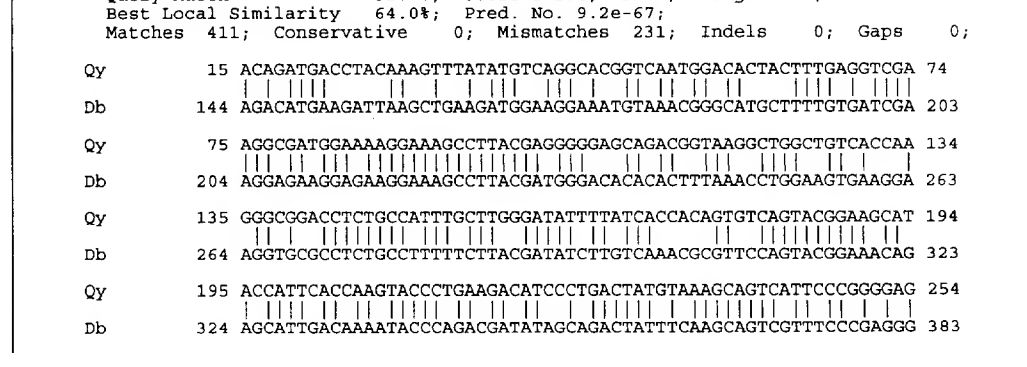
Query Match 43.1%; Score 362.6; DB 4; Length 699;
Best Local Similarity 73.1%; Pred. No. 4.7e-92;
Matches 480; Conservative 0; Mismatches 174; Indels 3; Gaps 1;

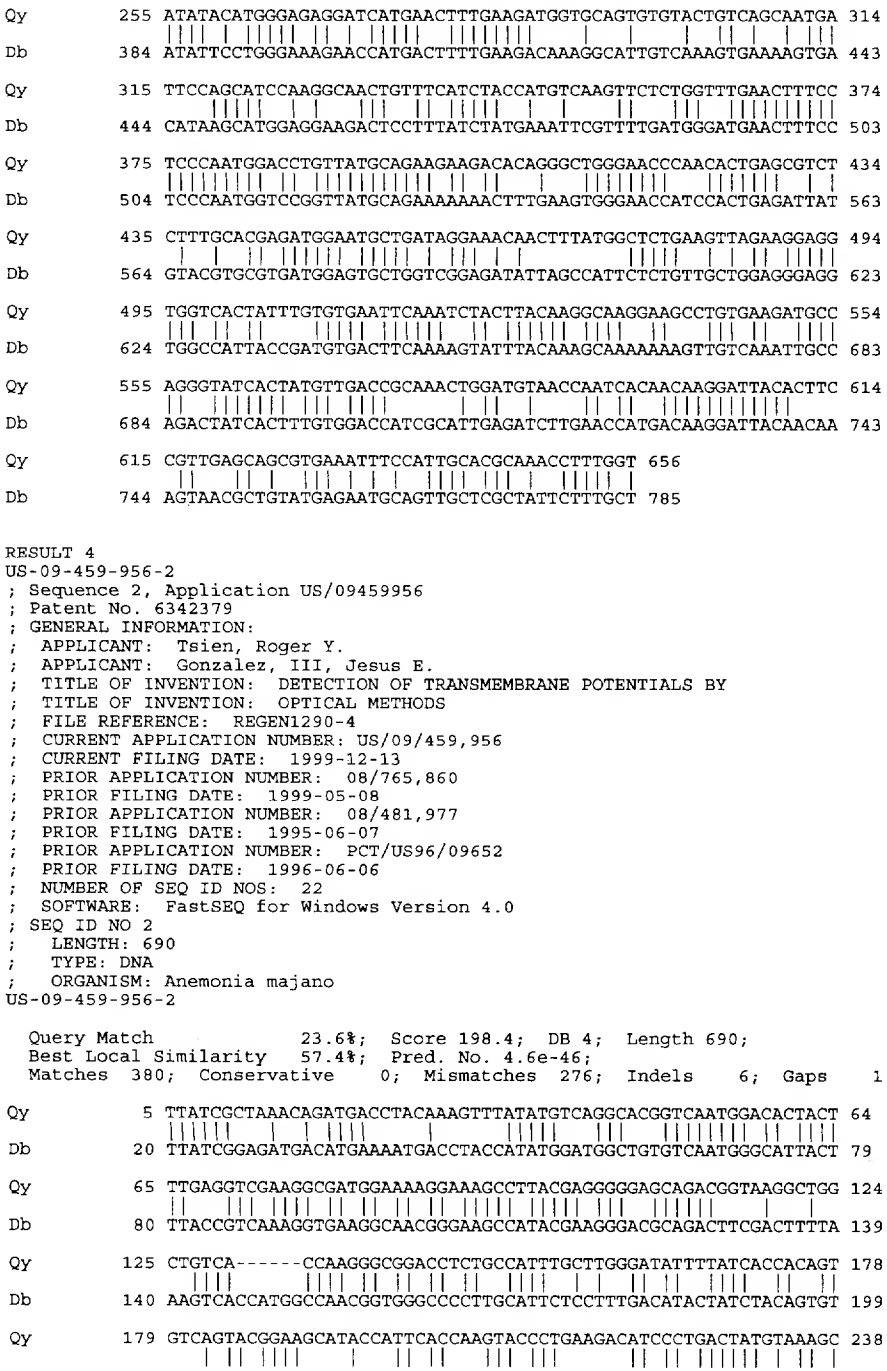




RESULT 3
US-09-459-956-7
; Sequence 7, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Clavularia sp
US-09-459-956-7

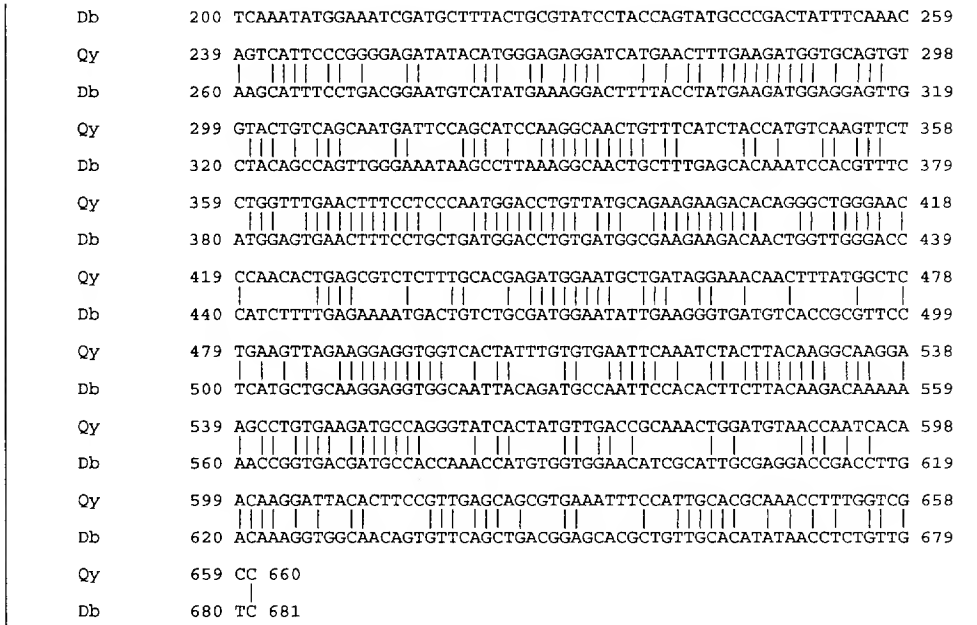
Query Match 32.4%; Score 272.4; DB 4; Length 801;
Best Local Similarity 64.0%; Pred. No. 9.2e-67;
Matches 411; Conservative 0; Mismatches 231; Indels 0; Gaps 0;





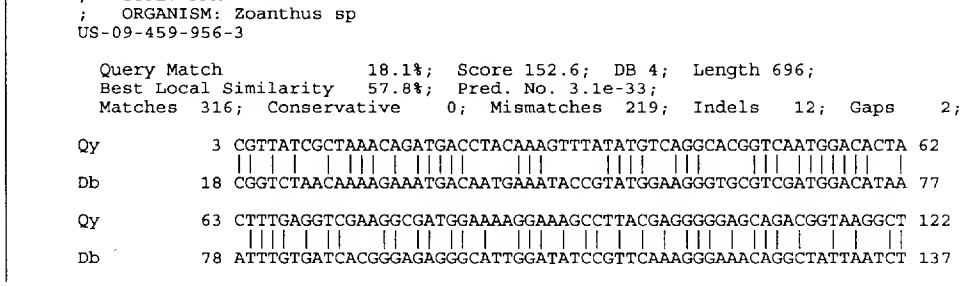
RESULT 4
US-09-459-956-2
; Sequence 2, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 690
; TYPE: DNA
; ORGANISM: Anemonia majano
US-09-459-956-2

Query Match 23.6%; Score 198.4; DB 4; Length 690;
Best Local Similarity 57.4%; Pred. No. 4.6e-46;
Matches 380; Conservative 0; Mismatches 276; Indels 6; Gaps 1;



RESULT 5
US-09-459-956-3
; Sequence 3, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Zoanthus sp
US-09-459-956-3

Query Match 18.1%; Score 152.6; DB 4; Length 696;
Best Local Similarity 57.8%; Pred. No. 3.1e-33;
Matches 316; Conservative 0; Mismatches 219; Indels 12; Gaps 2;



```
Qy        123 GGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCA 182
Db        138 GTGTGTGGTCGAAGGTGGACCATTGCCATTTGCCGAAGACATATTGTCAGCTGCCTTTAA 197

Qy        183 GTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTC 242
Db        198 CTACGGAAACAGGGTTTTCACTGAATATCCTCAAGACATAGTTGACTATTTCAAGAACTC 257

Qy        243 ATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAG------T 296
Db        258 GTGTCCTGCTGGATATACATGGGACAGGTCTTTTCTCTTTGAGGATGGAGCAGTTTGCAT 317

Qy        297 GTGTACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTT 356
Db        318 ATGTAATGCAGATATAACAGTGAGTGTTGAAGAAAACTGCATGTATCATGAGTCCAAATT 377

Qy        357 CTCTGGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGA 416
Db        378 TTATGGAGTGAATTTTCCTGCTGATGGACCTGTGATGAAAAAGATGACAGATAACTGGGA 437

Qy        417 ACCCAACACTGAGCGTCTCTTTGCACGA------GATGGAATGCTGATAGGAAACAACTT 470
Db        438 GCCATCCTGCGAGAAGATCATACCAGTACCTAAGCAGGGGATATTGAAAGGGGATGTCTC 497

Qy        471 TATGGCTCTGAAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAA 530
Db        498 CATGTACCTCCTTCTGAAGGATGGTGGGCGTTTACGGTGCCAATTCGACACAGTTTACAA 557

Qy        531 GGCAAGG 537
Db        558 AGCAAAG 564
```

RESULT 6
US-09-459-956-4
; Sequence 4, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Zoanthus sp
US-09-459-956-4

Query Match 17.9%; Score 150.2; DB 4; Length 696;
Best Local Similarity 57.5%; Pred. No. 1.5e-32;
Matches 335; Conservative 0; Mismatches 233; Indels 15; Gaps 3;

```
Qy         14 AACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCG 73
Db         29 AAGAAATGACAATGAAATACCACATGGAAGGGTGCGTCAACGGACATAAATTTGTGATCA 88

Qy         74 AAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGGCTGGCTGTCACCA 133
Db         89 CGGGCGAAGGCATTGGATATCCGTTCAAAGGGAAACAGACTATTAATCTGTGTGTGATCG 148

Qy        134 AGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCAGTACGGAAGCA 193
```



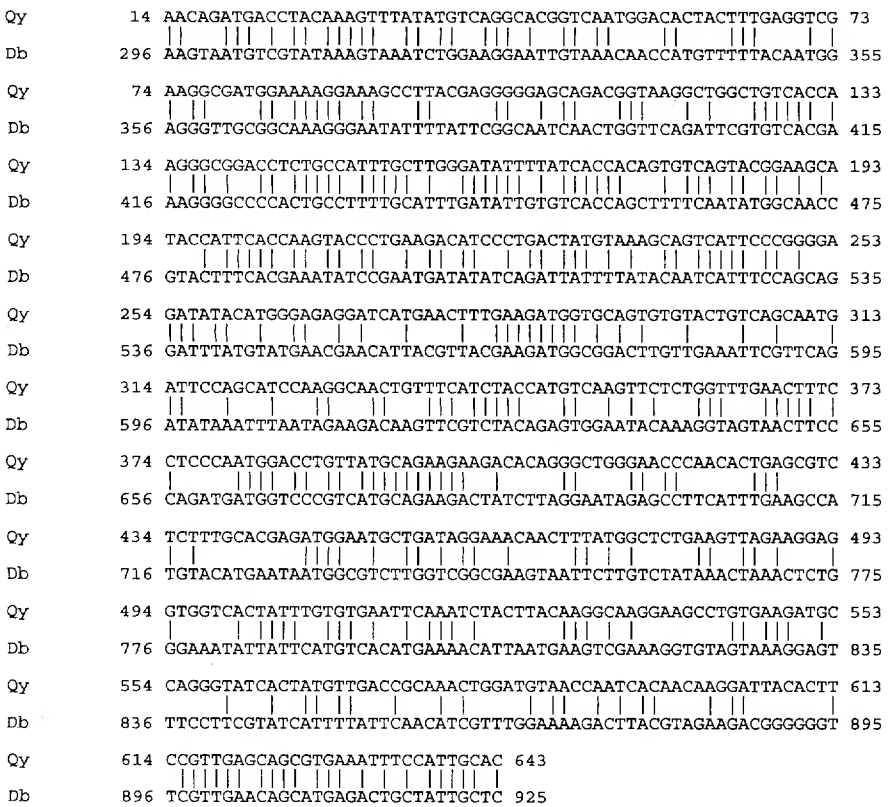
```
Db        149 AAGGGGGACCATTGCCATTTTCCGAAGACATATTGTCAGCTGGCTTTAAGTACGGAGACA 208

Qy        194 TACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCATTCCCGGGGA 253
Db        209 GGATTTTCACTGAATATCCTCAAGACATAGTAGACTATTTCAAGAACTCGTGTCCTGCTG 268

Qy        254 GATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCAATG 313
Db        269 GATATACATGGGGCAGGTCTTTTCTCTTTGAGGATGGAGCAGTCTGCATATGCAATGTAG 328

Qy        314 AT------TCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTGA 367
Db        329 ATATAACAGTGAGTGTCAAAGAAAACTGCATTTATCATAAGAGCATATTTAATGGAATGA 388

Qy        368 ACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAACACTG 427
Db        389 ATTTTCCTGCTGATGGACCTGTGATGAAAAAGATGACAACTAACTGGGAAGCATCCTGCG 448

Qy        428 AGCGTCTCTTTGCACGA------GATGGAATGCTGATAGGAAACAACTTTATGGCTCTGA 481
Db        449 AGAAGATCATGCCAGTACCTAAGCAGGGGATACTGAAAGGGGATGTCTCCATGTACCTCC 508

Qy        482 AGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAGG---A 538
Db        509 TTCTGAAGGATGGTGGGCGTTACCGGTGCCAGTTCGACACAGTTTACAAAGCAAAGTCTG 568

Qy        539 AGCCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACT 581
Db        569 TGCCAAGTAAGATGCCGGAGTGGCACTTCATCCAGCATAAGCT 611
```

RESULT 7
US-09-839-650-1
; Sequence 1, Application US/09839650
; Patent No. 6645761
; GENERAL INFORMATION:
; APPLICANT: Stratagene
; TITLE OF INVENTION: Humanized Polynucleotide Sequence Encoding Renilla Mulleri Green
; Patent No. 6645761
; TITLE OF INVENTION: Fluorescent Protein
; FILE REFERENCE: 25436/1755
; CURRENT APPLICATION NUMBER: US/09/839,650
; CURRENT FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 720
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Humanized R. mulleri polynucleotide
; NAME/KEY: misc_feature
; LOCATION: (1)..(720)
; OTHER INFORMATION: Humanized DNA sequence
US-09-839-650-1

Query Match 15.6%; Score 131.6; DB 4; Length 720;
Best Local Similarity 50.6%; Pred. No. 2.4e-27;
Matches 317; Conservative 0; Mismatches 309; Indels 0; Gaps 0;

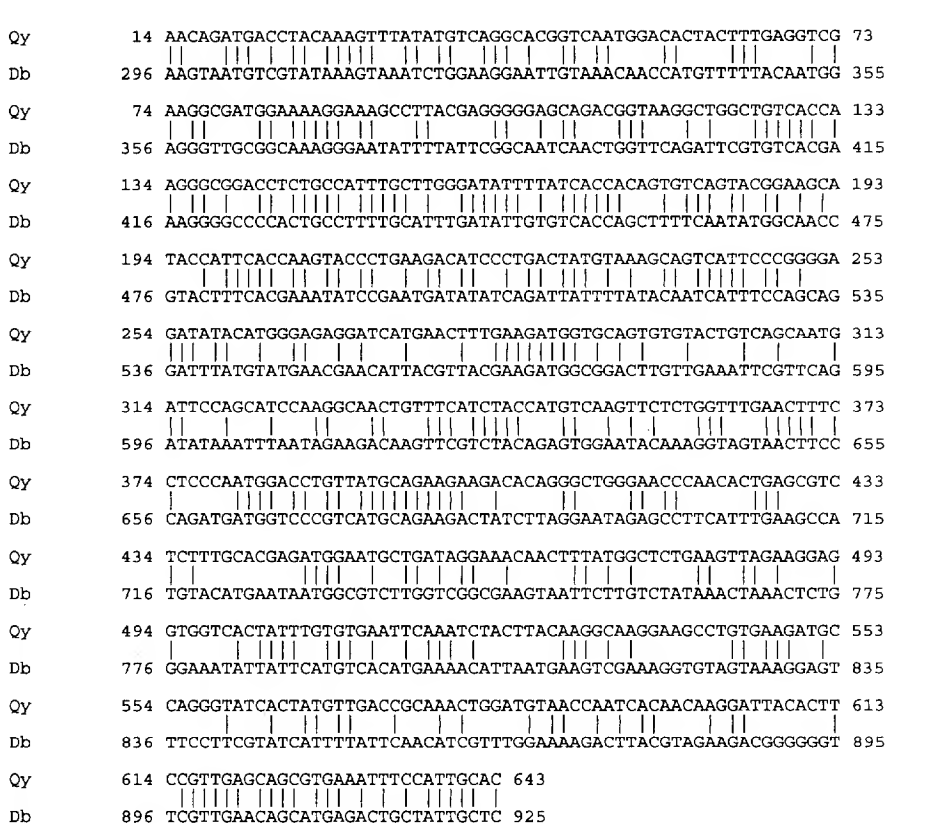
```
Qy         18 GATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCGAAGG 77
Db         45 GATGAGCTACAAGGTGAACCTGGAGGGCATCGTGAACAACCACGTGTTCACCATGGAGGG 104

Qy         78 CGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGGCTGGCTGTCACCAAGGG 137
Db        105 CTGCGGCAAGGGCAACATCCTGTTCGGCAACCAGCTGGTGCAGATCCGCGTGACCAAGGG 164

Qy        138 CGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCAGTACGGAAGCATACC 197
Db        165 CGCCCCCCTGCCCTTCGCCTTCGACATCGTGAGCCCCGCCTTCCAGTACGGCAACCGCAC 224

Qy        198 ATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCATTCCCGGGGAGATA 257
```

```
Db        225 CTTCACCAAGTACCCCAACGACATCAGCGACTACTTCATCCAGAGCTTCCCCGCCGGCTT 284

Qy        258 TACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCAATGATTC 317

Db        285 CATGTACGAGCGCACCCTGCGCTACGAGGACGGCGGCCTGGTGGAGATCCGCAGCGACAT 344

Qy        318 CAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTGAACTTTCCTCC 377

Db        345 CAACCTGATCGAGGACAAGTTCGTGTACCGCGTGGAGTACAAGGGCAGCAACTTCCCCGA 404

Qy        378 CAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAACACTGAGCGTCTCTT 437

Db        405 CGACGGCCCCGTGATGCAGAAGACCATCCTGGGCATCGAGCCCAGCTTCGAGGCCATGTA 464

Qy        438 TGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTGAAGTTAGAAGGAGGTGG 497

Db        465 CATGAACAACGGCGTGCTGGTGGGCGAGGTGATCCTGGTGTACAAGCTGAACAGCGGCAA 524

Qy        498 TCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAGGAAGCCTGTGAAGATGCCAGG 557

Db        525 GTACTACAGCTGCCACATGAAGACCCTGATGAAGAGCAAGGGCGTGGTGAAGGAGTTCCC 584

Qy        558 GTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAACAAGGATTACACTTCCGT 617

Db        585 CTCCTACCACTTCATCCAGCACCGCCTGGAGAAGACCTACGTGGAGGACGGCGGCTTCGT 644

Qy        618 TGAGCAGCGTGAAATTTCCATTGCAC 643

Db        645 GGAGCAGCACGAGACCGCCATCGCCC 670


RESULT 8
US-09-609-161B-15
; Sequence 15, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
;  APPLICANT: Bryan, Bruce
;  APPLICANT:  Szent-Gyorgyi, Christopher
;  APPLICANT:  PROLUME, LTD.
;  TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
;  TITLE OF INVENTION:  AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
;  TITLE OF INVENTION:  SCREENING AND NOVELTY ITEMS
;  FILE REFERENCE: 24729-121B
;  CURRENT APPLICATION NUMBER: US/09/609,161B
;  CURRENT FILING DATE:  2000-06-30
;  PRIOR APPLICATION NUMBER: 09/277,716
;  PRIOR FILING DATE: 1999-03-26
;  PRIOR APPLICATION NUMBER: 60/102,939
;  PRIOR FILING DATE: 1998-10-01
;  PRIOR APPLICATION NUMBER: 60/089,367
;  PRIOR FILING DATE: 1998-06-15
;  PRIOR APPLICATION NUMBER: 60/079,624
;  PRIOR FILING DATE: 1998-03-27
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
;   LENGTH: 1079
;   TYPE: DNA
;   ORGANISM: Renilla mulleri
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (259)..(975)
;   OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-609-161B-15

  Query Match           15.0%;  Score 126;  DB 4;  Length 1079;
  Best Local Similarity   50.0%;  Pred. No. 1.1e-25;
  Matches  315;  Conservative    0;  Mismatches  315;  Indels    0;  Gaps    0;

Qy         14 AACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCG 73

Db        296 AAGTAATGTCGTATAAAGTAAATCTGGAAGGAATTGTAAACAACCATGTTTTTACAATGG 355

Qy         74 AAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGGCTGGCTGTCACCA 133

Db        356 AGGGTTGCGGCAAAGGGAATATTTTATTCGGCAATCAACTGGTTCAGATTCGTGTCACGA 415

Qy        134 AGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCAGTACGGAAGCA 193

Db        416 AAGGGGCCCCACTGCCTTTTGCATTTGATATTGTGTCACCAGCTTTTCAATATGGCAACC 475

Qy        194 TACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCATTCCCGGGGA 253

Db        476 GTACTTTCACGAAATATCCGAATGATATATCAGATTATTTTATACAATCATTTCCAGCAG 535

Qy        254 GATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCAATG 313

Db        536 GATTTATGTATGAACGAACATTACGTTACGAAGATGGCGGACTTGTTGAAATTCGTTCAG 595

Qy        314 ATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTGAACTTTC 373

Db        596 ATATAAATTTAATAGAAGACAAGTTCGTCTACAGAGTGGAATACAAAGGTAGTAACTTCC 655

Qy        374 CTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAACACTGAGCGTC 433

Db        656 CAGATGATGGTCCCGTCATGCAGAAGACTATCTTAGGAATAGAGCCTTCATTTGAAGCCA 715

Qy        434 TCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTGAAGTTAGAAGGAG 493

Db        716 TGTACATGAATAATGGCGTCTTGGTCGGCGAAGTAATTCTTGTCTATAAACTAAACTCTG 775

Qy        494 GTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAGGAAGCCTGTGAAGATGC 553

Db        776 GGAAATATTATTCATGTCACATGAAAACATTAATGAAGTCGAAAGGTGTAGTAAAGGAGT 835

Qy        554 CAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAACAAGGATTACACTT 613

Db        836 TTCCTTCGTATCATTTTATTCAACATCGTTTGGAAAAGACTTACGTAGAAGACGGGGGGT 895

Qy        614 CCGTTGAGCAGCGTGAAATTTCCATTGCAC 643

Db        896 TCGTTGAACAGCATGAGACTGCTATTGCTC 925


RESULT 9
US-09-626-581D-64
; Sequence 64, Application US/09626581D
; Patent No. 6548249
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David
;  TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
;  TITLE OF INVENTION:  Libraries
;  FILE REFERENCE: A-66900-3/RMS
;  CURRENT APPLICATION NUMBER: US/09/626,581D
;  CURRENT FILING DATE:  2000-07-27
;  PRIOR APPLICATION NUMBER: 09/169,015
;  PRIOR FILING DATE: 1998-10-08
;  PRIOR APPLICATION NUMBER: 09/415,765
;  PRIOR FILING DATE: 1999-10-08
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
;   LENGTH: 1079
;   TYPE: DNA
;   ORGANISM: Renilla muelleri
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (259)..(975)
;   OTHER INFORMATION:
US-09-626-581D-64

  Query Match           15.0%;  Score 126;  DB 4;  Length 1079;
  Best Local Similarity   50.0%;  Pred. No. 1.1e-25;
  Matches  315;  Conservative    0;  Mismatches  315;  Indels    0;  Gaps    0;
```



RESULT 10
US-09-415-765B-64
; Sequence 64, Application US/09415765B
; Patent No. 6548632
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; TITLE OF INVENTION: Libraries
; FILE REFERENCE: A66900-1/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/415,765B
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Renilla muelleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; OTHER INFORMATION:
US-09-415-765B-64

Query Match 15.0%; Score 126; DB 4; Length 1079;
Best Local Similarity 50.0%; Pred. No. 1.1e-25;
Matches 315; Conservative 0; Mismatches 315; Indels 0; Gaps 0;



RESULT 11
US-09-626-580C-64
; Sequence 64, Application US/09626580C
; Patent No. 6562617
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Bogenberger, Jakob M.
; APPLICANT: Peele, Beau R.
; TITLE OF INVENTION: FUSIONS OF SCAFFOLD PROTEINS WITH RANDOM PEPTIDE LIBRARIES
; FILE REFERENCE: A-66900-2/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/626,580C
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 09/415,765
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Renilla muelleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; OTHER INFORMATION:
US-09-626-580C-64

```
  Query Match             15.0%;  Score 126;  DB 4;  Length 1079;
  Best Local Similarity   50.0%;  Pred. No. 1.1e-25;
  Matches  315;  Conservative    0;  Mismatches  315;  Indels    0;  Gaps    0;

Qy         14 AACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCG 73
              ||   ||| | || |||||  || ||  ||| |  || ||     ||     |||     | |
Db        296 AAGTAATGTCGTATAAAGTAAATCTGGAAGGAATTGTAAACAACCATGTTTTTACAATGG 355

Qy         74 AAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGGCTGGCTGTCACCA 133
              | ||     || ||||| ||    ||      || | ||   |||   | |   ||||| |
Db        356 AGGGTTGCGGCAAAGGGAATATTTTATTCGGCAATCAACTGGTTCAGATTCGTGTCACGA 415

Qy        134 AGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCAGTACGGAAGCA 193
              | || |  || ||||| ||||| |  |||||| | ||||||   | ||| || || | |
Db        416 AAGGGGCCCCACTGCCTTTTGCATTTGATATTGTGTCACCAGCTTTTCAATATGGCAACC 475

Qy        194 TACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCATTCCCGGGGA 253
              | ||||| || || ||  | || ||  | || ||| | |  || |||||  || |
Db        476 GTACTTTCACGAAATATCCGAATGATATATCAGATTATTTTATACAATCATTTCCAGCAG 535

Qy        254 GATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCAATG 313
              ||| ||  |  ||  | |    |  |||||||| | | |   |     | |     |
Db        536 GATTTATGTATGAACGAACATTACGTTACGAAGATGGCGGACTTGTTGAAATTCGTTCAG 595

Qy        314 ATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTGAACTTTC 373
              ||   |  |  ||  ||  ||| |||||  ||  | | |  |||  |||||  |
Db        596 ATATAAATTTAATAGAAGACAAGTTCGTCTACAGAGTGGAATACAAAGGTAGTAACTTCC 655

Qy        374 CTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAACACTGAGCGTC 433
              |    |||| || || ||||||||||  |    ||  || ||     |||
Db        656 CAGATGATGGTCCCGTCATGCAGAAGACTATCTTAGGAATAGAGCCTTCATTTGAAGCCA 715

Qy        434 TCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTGAAGTTAGAAGGAG 493
              | |          |||| |  || | ||  |     || | |     ||  || |    |
Db        716 TGTACATGAATAATGGCGTCTTGGTCGGCGAAGTAATTCTTGTCTATAAACTAAACTCTG 775

Qy        494 GTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAGGAAGCCTGTGAAGATGC 553
              |    | ||||  ||| |  | ||| |       ||| | |       || ||| |
Db        776 GGAAATATTATTCATGTCACATGAAAACATTAATGAAGTCGAAAGGTGTAGTAAAGGAGT 835

Qy        554 CAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAACAAGGATTACACTT 613
              |   |  || ||   |  | |     | ||  | | ||    | ||       |
Db        836 TTCCTTCGTATCATTTTATTCAACATCGTTTGGAAAAGACTTACGTAGAAGACGGGGGGT 895

Qy        614 CCGTTGAGCAGCGTGAAATTTCCATTGCAC 643
              |||||| |||| ||| | | | |||||
Db        896 TCGTTGAACAGCATGAGACTGCTATTGCTC 925


RESULT 12
US-09-277-716-15
; Sequence 15, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
;  APPLICANT: Bryan, Bruce
;  APPLICANT: Szent-Gyorgyi, Christopher
;  APPLICANT: PROLUME, LTD.
;  TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE   L
;  CURRENT APPLICATION NUMBER: US/09/277,716A
;  CURRENT FILING DATE: 1999-03-26
;  EARLIER APPLICATION NUMBER: 60/102,939
;  EARLIER FILING DATE: 1998-10-01
;  EARLIER APPLICATION NUMBER: 60/089,367
;  EARLIER FILING DATE: 1998-06-15
;  EARLIER APPLICATION NUMBER: 60/079,624
;  EARLIER FILING DATE: 1998-03-27
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
;   LENGTH: 1079
;   TYPE: DNA
;   ORGANISM: Renilla mulleri
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (259)..(975)
;   FEATURE:
;   OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-277-716-15

  Query Match             15.0%;  Score 126;  DB 3;  Length 1085;
  Best Local Similarity   50.0%;  Pred. No. 1.1e-25;
  Matches  315;  Conservative    0;  Mismatches  315;  Indels    0;  Gaps    0;

Qy         14 AACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCG 73
              ||   ||| | || |||||  || ||  ||| |  || ||     ||     |||     | |
Db        296 AAGTAATGTCGTATAAAGTAAATCTGGAAGGAATTGTAAACAACCATGTTTTTACAATGG 355

Qy         74 AAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGGCTGGCTGTCACCA 133
              | ||     || ||||| ||    ||      || | ||   |||   | |   ||||| |
Db        356 AGGGTTGCGGCAAAGGGAATATTTTATTCGGCAATCAACTGGTTCAGATTCGTGTCACGA 415

Qy        134 AGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCAGTACGGAAGCA 193
              | || |  || ||||| ||||| |  |||||| | ||||||   | ||| || || | |
Db        416 AAGGGGCCCCACTGCCTTTTGCATTTGATATTGTGTCACCAGCTTTTCAATATGGCAACC 475

Qy        194 TACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCATTCCCGGGGA 253
              | ||||| || || ||  | || ||  | || ||| | |  || |||||  || |
Db        476 GTACTTTCACGAAATATCCGAATGATATATCAGATTATTTTATACAATCATTTCCAGCAG 535

Qy        254 GATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCAATG 313
              ||| ||  |  ||  | |    |  |||||||| | | |   |     | |     |
Db        536 GATTTATGTATGAACGAACATTACGTTACGAAGATGGCGGACTTGTTGAAATTCGTTCAG 595

Qy        314 ATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTGAACTTTC 373
              ||   |  |  ||  ||  ||| |||||  ||  | | |  |||  |||||  |
Db        596 ATATAAATTTAATAGAAGACAAGTTCGTCTACAGAGTGGAATACAAAGGTAGTAACTTCC 655

Qy        374 CTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAACACTGAGCGTC 433
              |    |||| || || ||||||||||  |    ||  || ||     |||
Db        656 CAGATGATGGTCCCGTCATGCAGAAGACTATCTTAGGAATAGAGCCTTCATTTGAAGCCA 715

Qy        434 TCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTGAAGTTAGAAGGAG 493
              | |          |||| |  || | ||  |     || | |     ||  || |    |
Db        716 TGTACATGAATAATGGCGTCTTGGTCGGCGAAGTAATTCTTGTCTATAAACTAAACTCTG 775

Qy        494 GTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAGGAAGCCTGTGAAGATGC 553
              |    | ||||  ||| |  | ||| |       ||| | |       || ||| |
Db        776 GGAAATATTATTCATGTCACATGAAAACATTAATGAAGTCGAAAGGTGTAGTAAAGGAGT 835

Qy        554 CAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAACAAGGATTACACTT 613
              |   |  || ||   |  | |     | ||  | | ||    | ||       |
Db        836 TTCCTTCGTATCATTTTATTCAACATCGTTTGGAAAAGACTTACGTAGAAGACGGGGGGT 895

Qy        614 CCGTTGAGCAGCGTGAAATTTCCATTGCAC 643
              |||||| |||| ||| | | | |||||
Db        896 TCGTTGAACAGCATGAGACTGCTATTGCTC 925


RESULT 13
US-09-839-650-2
; Sequence 2, Application US/09839650
; Patent No. 6645761
; GENERAL INFORMATION:
;  APPLICANT: Stratagene
;  TITLE OF INVENTION: Humanized Polynucleotide Sequence Encoding Renilla Mulleri Green
; Patent No. 6645761
;  TITLE OF INVENTION:  Fluorescent Protein
;  FILE REFERENCE: 25436/1755
;  CURRENT APPLICATION NUMBER: US/09/839,650
;  CURRENT FILING DATE:  2001-04-19
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 1021
;   TYPE: DNA
```

; ORGANISM: Renilla muelleri
; FEATURE:
; NAME/KEY: exon
; LOCATION: (259)..(976)
US-09-839-650-2

Query Match 14.6%; Score 122.8; DB 4; Length 1021;
Best Local Similarity 49.7%; Pred. No. 8.2e-25;
Matches 313; Conservative 0; Mismatches 317; Indels 0; Gaps 0;

```
Qy        14 AACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCG 73
             ||   ||| | || |||||  || ||  ||| |  || ||    ||     |||    | |
Db       296 AAGTAATGTCGTATAAAGTAAATCTGGAAGGAATTGTAAACAACCATGTTTTTACAATGG 355

Qy        74 AAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGGCTGGCTGTCACCA 133
             | ||     | ||| | ||   ||     ||  | ||   |||     |  ||||| |
Db       356 AGGGTTGCCGCAAACGGAATATTTTATTCGGCAATCAACTGGTTCACATTCGTGTCACGA 415

Qy       134 AGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCAGTACGGAAGCA 193
             | || || || ||||| ||||| |  |||||| | |||||    | ||| || || | |
Db       416 AAGGGGGCCCACTGCCTTTTGCATTTGATATTGTGTCACCAGCTTTTCAATATGGCAACC 475

Qy       194 TACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCATTCCCGGGGA 253
               | ||||| || || ||  | || ||  | || ||| | |  || ||||| || |
Db       476 GTACTTTCACGAAATATCCGAATGATATATCAGATTATTTTATACAATCATTTCCAGCAG 535

Qy       254 GATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCAATG 313
             ||| ||  |  ||  | |   |   |  |||||||| | | |   |  |  |   |
Db       536 GATTTATGTATGAACGAACATTACGTTACGAAGATGGCGGACTTGTTGAAATTCGTTCAG 595

Qy       314 ATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTGAACTTTC 373
             ||   |   |  ||  ||  ||| |||||  ||  | | |  |||  |||||  |
Db       596 ATATAAATTTAATAGAAGACAAGTTCGTCTACAGAGTGGAATACAAAGGTAGTAACTTCC 655

Qy       374 CTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAACACTGAGCGTC 433
             |    |||| || || ||||||||||  |    ||   || ||     |||
Db       656 CAGATGATGGTCCCGTCATGCAGAAGACTATCTTAGGAATAGAGCCTTCATTTGAAGCCA 715

Qy       434 TCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTGAAGTTAGAAGGAG 493
             | |      ||||  |  || | ||  |    || | |    ||  || | |  |
Db       716 TGTACATGAATAATGGCGTCTTGGTCGGCGAAGTAATTCTTGTCTATAAACTAAACTCTG 775

Qy       494 GTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAGGAAGCCTGTGAAGATGC 553
             |    | ||||  ||| |  | ||| |       ||| | |          || ||| |
Db       776 GGAAATATTATTCATGTCACATGAAAACATTAATGAAGTCGAAAGGTGTAGTAAAGGAGT 835

Qy       554 CAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAACAAGGATTACACTT 613
               |  |  ||  ||   |  | |     | ||  | | ||   | ||       |
Db       836 TTCCTTCGTATCATTTTATTCAACATCGTTTGGAAAAGACTTACGTAGAAGACGGGGGGT 895

Qy       614 CCGTTGAGCAGCGTGAAATTTCCATTGCAC 643
              |||||| |||| ||| | | | ||||| |
Db       896 TCGTTGAACAGCATGAGACTGCTATTGCTC 925
```

RESULT 14
US-09-277-716-30
; Sequence 30, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34)..(747)
; FEATURE:
; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert A)
US-09-277-716-30

Query Match 14.6%; Score 122.6; DB 3; Length 1104;
Best Local Similarity 49.8%; Pred. No. 9.7e-25;
Matches 311; Conservative 0; Mismatches 314; Indels 0; Gaps 0;

```
Qy        19 ATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCGAAGGC 78
             |||  |    |||| |    |  ||| |  || ||    |||   |||     | |||||
Db        76 ATGTCGGCAAAAGCTAGCGTTGAAGGAATCGTGAACAATCACGTTTTTTCCATGGAAGGA 135

Qy        79 GATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGGCTGGCTGTCACCAAGGGC 138
             |||||||||  ||    |    ||  | ||  | |      |    || ||  |||||
Db       136 TTTGGAAAAGGCAATGTATTATTTGGAAACCAATTGATGCAAATCCGGGTTACAAAGGGA 195

Qy       139 GGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCAGTACGGAAGCATACCA 198
             || ||  ||||||| ||||  ||||||  | ||   |    |  || |||||  |    |
Db       196 GGTCCGTTGCCATTCGCTTTCGATATTGTTTCCATAGCTTTCCAATACGGGAATCGCACT 255

Qy       199 TTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCATTCCCGGGGAGATAT 258
             |||||  || |||||  || |||||   | |||||   |     || ||||||||||    ||| |
Db       256 TTCACGAAATACCCAGACGACATTGCGGACTACTTTGTTCAATCATTCCCGGCTGGATTT 315

Qy       259 ACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCAATGATTCC 318
               |  || || |   |   ||||||||||||  ||  |   |   |   |   |||
Db       316 TTCTACGAAAGAAATCTACGCTTTGAAGATGGCGCCATTGTTGACATTCGTTCAGATATA 375

Qy       319 AGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTGAACTTTCCTCCC 378
             ||  |  |||   |   |||  |||| | ||  |||     ||      || |||
Db       376 AGTTTAGAAGATGATAAGTTCCACTACAAAGTGGAGTATAGAGGCAACGGTTTCCCTAGT 435

Qy       379 AATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAACACTGAGCGTCTCTTT 438
             || ||||| || ||||| ||     |  |||  |||  ||       ||||    |||
Db       436 AACGGACCCGTGATGCAAAAAGCCATCCTCGGCATGGAGCCATCGTTTGAGGTGGTCTAC 495

Qy       439 GCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTGAAGTTAGAAGGAGGTGGT 498
               ||  | ||| | ||  |       | | |  |  ||  | ||    |||
Db       496 ATGAACAGCGGCGTTCTGGTGGGCGAAGTAGATCTCGTTTACAAACTCGAGTCAGGGAAC 555

Qy       499 CACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAGGAAGCCTGTGAAGATGCCAGGG 558
             | || | |||  |  | ||| |  |||||  | |          |||||         |
Db       556 TATTACTCGTGCCACATGAAAACGTTTTACAGATCCAAAGGTGGAGTGAAAGAATTCCCG 615

Qy       559 TATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAACAAGGATTACACTTCCGTT 618
               |  |  |  ||  ||   | |     |  |  |   ||    | ||    |  | |||
Db       616 GAATATCACTTTATCCATCATCGTCTGGAGAAAACCTACGTGGAAGAAGGAAGCTTCGTG 675

Qy       619 GAGCAGCGTGAAATTTCCATTGCAC 643
             || ||  |  ||  |   |||||||||
Db       676 GAACAACACGAGACGGCCATTGCAC 700
```

RESULT 15
US-09-609-161B-30
; Sequence 30, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B

```
;  CURRENT APPLICATION NUMBER: US/09/609,161B
;  CURRENT FILING DATE:  2000-06-30
;  PRIOR APPLICATION NUMBER: 09/277,716
;  PRIOR FILING DATE: 1999-03-26
;  PRIOR APPLICATION NUMBER: 60/102,939
;  PRIOR FILING DATE: 1998-10-01
;  PRIOR APPLICATION NUMBER: 60/089,367
;  PRIOR FILING DATE: 1998-06-15
;  PRIOR APPLICATION NUMBER: 60/079,624
;  PRIOR FILING DATE: 1998-03-27
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
;   LENGTH: 1104
;   TYPE: DNA
;   ORGANISM: Ptilosarcus gurneyi
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (34)..(747)
;   OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert A)
US-09-609-161B-30

  Query Match             14.6%;  Score 122.6;  DB 4;  Length 1104;
  Best Local Similarity   49.8%;  Pred. No. 9.7e-25;
  Matches  311;  Conservative    0;  Mismatches  314;  Indels    0;  Gaps    0;

Qy         19 ATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCGAAGGC 78
              ||| |    |||| |    |   ||| |  || ||    |||   |||    | |||||
Db         76 ATGTCGGCAAAAGCTAGCGTTGAAGGAATCGTGAACAATCACGTTTTTTCCATGGAAGGA 135

Qy         79 GATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGGCTGGCTGTCACCAAGGGC 138
              |||||||| ||     |      ||  | ||   | |      |     || || |||||
Db        136 TTTGGAAAAGGCAATGTATTATTTGGAAACCAATTGATGCAAATCCGGGTTACAAAGGGA 195

Qy        139 GGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCAGTACGGAAGCATACCA 198
              || ||  ||||||| ||||  |||||| | ||   |   |  || ||||| |     |
Db        196 GGTCCGTTGCCATTCGCTTTCGATATTGTTTCCATAGCTTTCCAATACGGGAATCGCACT 255

Qy        199 TTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCATTCCCGGGGAGATAT 258
              ||||| || |||||  || |||||  | |||||  |     || ||||||||||   ||| |
Db        256 TTCACGAAATACCCAGACGACATTGCGGACTACTTTGTTCAATCATTCCCGGCTGGATTT 315

Qy        259 ACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCAATGATTCC 318
                |  || || |   |   ||||||||||||| ||  |   |     |  |    |||
Db        316 TTCTACGAAAGAAATCTACGCTTTGAAGATGGCGCCATTGTTGACATTCGTTCAGATATA 375

Qy        319 AGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTGAACTTTCCTCCC 378
              ||  |  |||   |   |||  |||| | ||  |||     ||       || |||
Db        376 AGTTTAGAAGATGATAAGTTCCACTACAAAGTGGAGTATAGAGGCAACGGTTTCCCTAGT 435

Qy        379 AATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAACACTGAGCGTCTCTTT 438
              || ||||| || ||||| ||    |  |  |||  ||| ||      ||||    |||
Db        436 AACGGACCCGTGATGCAAAAAGCCATCCTCGGCATGGAGCCATCGTTTGAGGTGGTCTAC 495

Qy        439 GCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTGAAGTTAGAAGGAGGTGGT 498
                  ||  | ||| | ||  |       | | | |   ||  | ||   |||
Db        496 ATGAACAGCGGCGTTCTGGTGGGCGAAGTAGATCTCGTTTACAAACTCGAGTCAGGGAAC 555

Qy        499 CACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAGGAAGCCTGTGAAGATGCCAGGG 558
              | || | |||  |  | ||| |   |||||   | |       |||||        |
Db        556 TATTACTCGTGCCACATGAAAACGTTTTACAGATCCAAAGGTGGAGTGAAAGAATTCCCG 615

Qy        559 TATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAACAAGGATTACACTTCCGTT 618
              |  |  |  ||  ||   | |     |  |  |   ||    | ||    |  | |||
Db        616 GAATATCACTTTATCCATCATCGTCTGGAGAAAACCTACGTGGAAGAAGGAAGCTTCGTG 675

Qy        619 GAGCAGCGTGAAATTTCCATTGCAC 643
              || || |  || |   |||||||||
Db        676 GAACAACACGAGACGGCCATTGCAC 700


Search completed: August 13, 2004, 20:48:15
Job time : 79 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 13, 2004, 11:19:50 ; Search time 3675 Seconds
(without alignments)
9918.773 Million cell updates/sec

Title: US-09-890-463-6
Perfect score: 841
Sequence: 1 tccgttatcgctaaacagat.....aaaagcggccgctcgaatta 841

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	841	100.0	841	6	BD248906	BD248906 Pigment p
2	824.6	98.0	841	6	BD248905	BD248905 Pigment p
3	710.4	84.5	881	3	AF383156	AF383156 Goniopora
4	657.8	78.2	693	6	AX699793	AX699793 Sequence
5	652.4	77.6	663	6	AX699755	AX699755 Sequence
6	648.8	77.1	660	6	AX699783	AX699783 Sequence
7	647.2	77.0	660	6	AX699753	AX699753 Sequence
8	647.2	77.0	660	6	AX699763	AX699763 Sequence
9	647.2	77.0	660	6	AX699785	AX699785 Sequence
10	647.2	77.0	669	6	AX699923	AX699923 Sequence
11	645.6	76.8	660	6	AX699759	AX699759 Sequence
12	642.4	76.4	669	6	AX699921	AX699921 Sequence
13	640.8	76.2	669	6	AX699927	AX699927 Sequence
14	640.8	76.2	669	6	AX699933	AX699933 Sequence
15	639.2	76.0	660	6	AX699781	AX699781 Sequence
16	639.2	76.0	669	6	AX699929	AX699929 Sequence
17	637.8	75.8	663	6	AX699871	AX699871 Sequence
18	637.8	75.8	663	6	AX699873	AX699873 Sequence
19	637.6	75.8	669	6	AX699925	AX699925 Sequence
20	636.2	75.6	663	6	AX699827	AX699827 Sequence
21	636.2	75.6	663	6	AX699829	AX699829 Sequence
22	636.2	75.6	765	6	AX699819	AX699819 Sequence
23	636.2	75.6	765	6	AX699821	AX699821 Sequence
24	636	75.6	660	6	AX699811	AX699811 Sequence
25	633.2	75.3	660	6	AX699823	AX699823 Sequence
26	631.4	75.1	660	6	AX699825	AX699825 Sequence
27	631.2	75.1	660	6	AX699771	AX699771 Sequence
28	631.2	75.1	660	6	AX699773	AX699773 Sequence
29	631.2	75.1	660	6	AX699777	AX699777 Sequence
30	631.2	75.1	660	6	AX699803	AX699803 Sequence
31	631.2	75.1	660	6	AX699815	AX699815 Sequence
32	631.2	75.1	660	6	AX699817	AX699817 Sequence
33	629.8	74.9	663	6	AX699835	AX699835 Sequence
34	629.8	74.9	663	6	AX699847	AX699847 Sequence
35	629.8	74.9	663	6	AX699893	AX699893 Sequence
36	629.6	74.9	660	6	AX699779	AX699779 Sequence
37	628.2	74.7	660	6	AX699899	AX699899 Sequence
38	628.2	74.7	663	6	AX699831	AX699831 Sequence
39	628.2	74.7	663	6	AX699833	AX699833 Sequence
40	628.2	74.7	663	6	AX699879	AX699879 Sequence
41	628	74.7	660	6	AX699775	AX699775 Sequence
42	628	74.7	660	6	AX699805	AX699805 Sequence
43	628	74.7	660	6	AX699809	AX699809 Sequence
44	628	74.7	660	6	AX699813	AX699813 Sequence
45	627.8	74.6	660	6	AX699769	AX699769 Sequence

ALIGNMENTS

RESULT 1
BD248906
LOCUS BD248906 841 bp DNA linear PAT 17-JUL-2003
DEFINITION Pigment protein from coral tissue.
ACCESSION BD248906
VERSION BD248906.1 GI:33058676
KEYWORDS JP 2002535978-A/2.
SOURCE Acropora aspera
ORGANISM Acropora aspera
Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
Astrocoeniina; Acroporidae; Acropora.
REFERENCE 1 (bases 1 to 841)
AUTHORS Guldberg,O.H. and Dove,S.
TITLE Pigment protein from coral tissue
JOURNAL Patent: JP 2002535978-A 2 29-OCT-2002;

```
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Acropora aspera (plate coral)
            PN   JP 2002535978-A/2
            PD   29-OCT-2002
            PF   02-FEB-2000 JP 2000597303
            PR   02-FEB-1999 AU      PP 8463
            PI   OVE HOEGH GULDBERG,SOPHIE DOVE
            PC   C12N15/09,A61K7/42,C07K14/435,C09B61/00,C09K11/06,C12N1/15, PC
              C12N1/19,
            PC
            C12N1/21,C12N5/10,C12P21/02,C12Q1/68,G01N21/78,C12N15/00,C12N5/ PC
             00
            CC   Pigment protein from coral tissue
            FH   Key             Location/Qualifiers
            FT   source          1. .841
            FT                   /organism='Acropora aspera (plate coral)'.
FEATURES              Location/Qualifiers
     source           1. .841
                      /organism="Acropora aspera"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:140238"
ORIGIN

  Query Match             100.0%;  Score 841;  DB 6;  Length 841;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-198;
  Matches  841;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60

Qy         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120

Qy        121 CTGGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGT 180

Qy        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240

Qy        241 TCATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300

Qy        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360

Qy        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420

Qy        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480

Qy        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 540

Qy        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600

Qy        601 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660

Qy        661 TGCCGTTTTTTCAGAGTCAAATCAAGGCACAAATACGCAGTGGCGTAAAAAACGTAGATT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TGCCGTTTTTTCAGAGTCAAATCAAGGCACAAATACGCAGTGGCGTAAAAAACGTAGATT 720

Qy        721 CTGATTTTAGCTTATAGAAGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAAACT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTGATTTTAGCTTATAGAAGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAAACT 780

Qy        781 TTTGAAAACAACGCCATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGGCCGCTCGAATT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTGAAAACAACGCCATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGGCCGCTCGAATT 840

Qy        841 A 841
              |
Db        841 A 841


RESULT 2
BD248905
LOCUS       BD248905                841 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Pigment protein from coral tissue.
ACCESSION   BD248905
VERSION     BD248905.1  GI:33058675
KEYWORDS    JP 2002535978-A/1.
SOURCE      Acropora aspera
  ORGANISM  Acropora aspera
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
            Astrocoeniina; Acroporidae; Acropora.
REFERENCE   1  (bases 1 to 841)
  AUTHORS   Guldberg,O.H. and Dove,S.
  TITLE     Pigment protein from coral tissue
  JOURNAL   Patent: JP 2002535978-A 1 29-OCT-2002;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Acropora aspera (plate coral)
            PN   JP 2002535978-A/1
            PD   29-OCT-2002
            PF   02-FEB-2000 JP 2000597303
            PR   02-FEB-1999 AU      PP 8463
            PI   OVE HOEGH GULDBERG,SOPHIE DOVE
            PC   C12N15/09,A61K7/42,C07K14/435,C09B61/00,C09K11/06,C12N1/15, PC
              C12N1/19,
            PC
            C12N1/21,C12N5/10,C12P21/02,C12Q1/68,G01N21/78,C12N15/00,C12N5/ PC
             00
            CC   Pigment protein from coral tissue
            FH   Key             Location/Qualifiers
            FT   source          1. .841
            FT                   /organism='Acropora aspera (plate coral)'.
FEATURES              Location/Qualifiers
     source           1. .841
                      /organism="Acropora aspera"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:140238"
ORIGIN

  Query Match             98.0%;  Score 824.6;  DB 6;  Length 841;
  Best Local Similarity   98.7%;  Pred. No. 6e-194;
  Matches  830;  Conservative    1;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60

Qy         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120

Qy        121 CTGGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGT 180

Qy        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240
```

```
Qy        241 TCATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300

Qy        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360

Qy        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420

Qy        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480

Qy        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAGGAAG 540

Qy        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600

Qy        601 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        601 AAGGATTACACTTCCGTTGAGCAGCGTGAAATTTCCATTGCACGCAAACCTTTGGTCGCC 660

Qy        661 TGCCGTTTTTTCAGAGTCAAATCAAGGCACAAATACGCAGTGGCGTAAAAAACGTAGATT 720
              ||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        661 TGCTGTTTTTTCAGAGTCAAATCAAGGCACAAATAAGCAGTGGCGTAAAAAACGTAGATT 720

Qy        721 CTGATTTTAGCTTATAGAAGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAAACT 780
              |||||||||||||| |||||||||||||||||||| |||||| | ||||||||||||||
Db        721 CTGATTTTAGCTTAGAGAAGTAGGAACGAAGAAGTGTAGACAACCTTCAATGATTAAACT 780

Qy        781 TTTGAAAACAACGCCATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGGCCGCTCGAATT 840
              ||||||||||||:||| ||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTGAAAACAACSCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGGCCGCTCGAATT 840

Qy        841 A 841
              |
Db        841 A 841
```

```
RESULT 3
AF383156
LOCUS       AF383156                881 bp    mRNA    linear   INV 05-NOV-2001
DEFINITION  Goniopora tenuidens GFP-like chromoprotein mRNA, complete cds.
ACCESSION   AF383156
VERSION     AF383156.1  GI:16660127
KEYWORDS    .
SOURCE      Goniopora tenuidens
  ORGANISM  Goniopora tenuidens
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
            Fungiina; Poritidae; Goniopora.
REFERENCE   1  (bases 1 to 881)
  AUTHORS   Gurskaya,N.G., Fradkov,A.F., Terskikh,A., Matz,M.V., Labas,Y.A.,
            Martynov,V.I., Yanushevich,Y.G., Lukyanov,K.A. and Lukyanov,S.A.
  TITLE     GFP-like chromoproteins as a source of far-red fluorescent proteins
  JOURNAL   FEBS Lett. 507 (1), 16-20 (2001)
  MEDLINE   21538626
   PUBMED   11682051
REFERENCE   2  (bases 1 to 881)
  AUTHORS   Gurskaya,N.G., Lukyanov,K.A., Labas,Y.A. and Lukyanov,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2001) Institute of Bioorganic Chemistry,
            Miklukho-Maklaya 16/10, Moscow 117997, Russia
FEATURES             Location/Qualifiers
     source          1. .881
                     /organism="Goniopora tenuidens"
                     /mol_type="mRNA"
                     /db_xref="taxon:75301"
     CDS             83. .748
                     /note="gtCP"
                     /codon_start=1
                     /product="GFP-like chromoprotein"
                     /protein_id="AAL27542.1"
                     /db_xref="GI:16660128"
                     /translation="MSVIAKQMTYKVYMSGTVNGHYFEVQGDGKGKPYEGEQTVKLTV
                     TKGGPLPFAWDILSPQSQYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCT
                     VSNDSSIQGNCFIYNVKFSGLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMA
                     LKLEGGGHYLCEFKSTYKAKKPVKMPGYHYVDRKLDVTNHNIDYTSVEQCEISIARKP
                     VVA"
ORIGIN
```

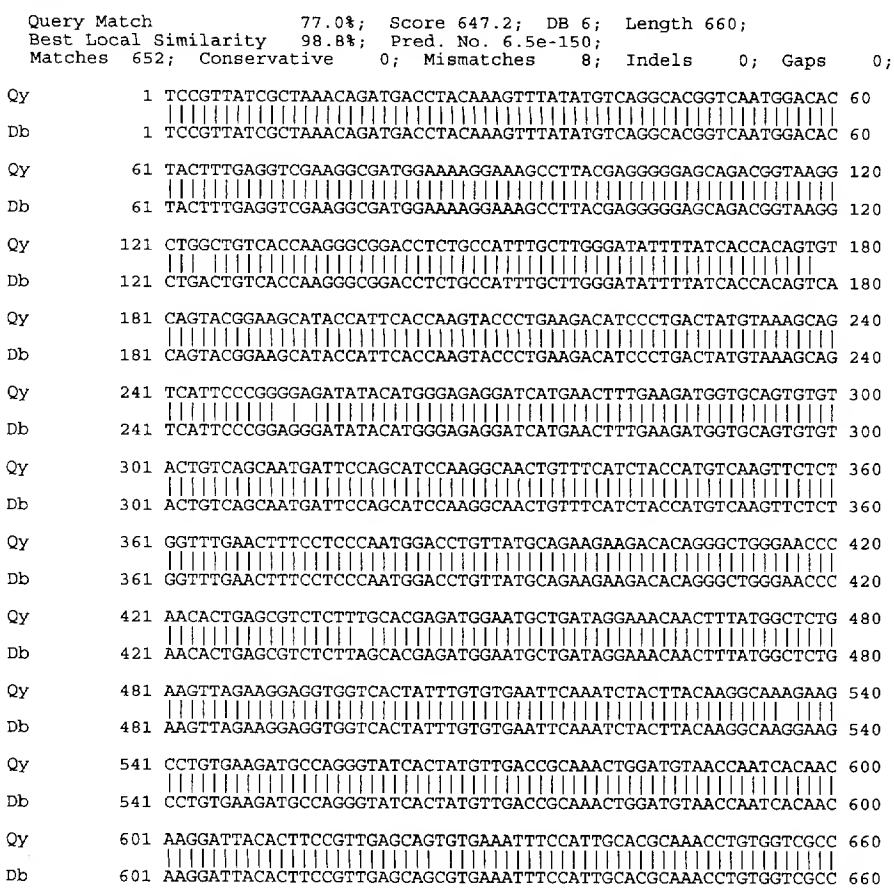
```
  Query Match             84.5%;  Score 710.4;  DB 3;  Length 881;
  Best Local Similarity   95.4%;  Pred. No. 1.3e-165;
  Matches  754;  Conservative    0;  Mismatches   31;  Indels    5;  Gaps    2;

Qy          4 GTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACACTAC 63
              || |||||||||| |||||||||| ||||||||||||||||||||||||||||||||||
Db         89 GTGATCGCTAAACAAATGACCTACAAGGTTTATATGTCAGGCACGGTCAATGGACACTAC 148

Qy         64 TTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGGCTG 123
              |||||||| |||||||||||||||||||||||||||||||||||||||||||| |||
Db        149 TTTGAGGTCCAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAAGCTC 208

Qy        124 GCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGTCAG 183
              |||||||||||| |||||||||||||||||||||||||||||||||||||||| ||||
Db        209 ACTGTCACCAAGGGTGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTCTCAG 268

Qy        184 TACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        269 TACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCA 328

Qy        244 TTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACT 303
              ||||| | | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        329 TTCCCTGAGGGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACT 388

Qy        304 GTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGT 363
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        389 GTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACAATGTCAAGTTCTCTGGT 448

Qy        364 TTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAAC 423
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        449 TTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAAC 508

Qy        424 ACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTGAAG 483
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        509 ACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTGAAG 568

Qy        484 TTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAGCCT 543
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        569 TTGGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAGCCT 628

Qy        544 GTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAACAAG 603
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        629 GTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAACATA 688

Qy        604 GATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCCTGC 663
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        689 GATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCCTGA 748

Qy        664 CGTTTTTTCAGAGTCAAATCAAGGCACA----AATACGCAGTGGCGTAAAAAACGTAGAT 719
              |||||||||||||||||||||||||||| |    |||||||||||||||| ||||||||||
Db        749 CGTTTTTTCAGAGTCAAATCAAGGCAGAAAATAATACGCAGTGGCGTAAGAAACGTAGAT 808

Qy        720 TCTGATTTTAGCTTATAGAAGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAAAC 779
              | |||||||||||| |||||||||||||||  ||||||||||||   |||||||||
Db        809 TTTGATTTTAGCTTAGAGAAGTAGGAACGAA-CAGTGTAAACAACCCCGAATGATTAACT 867

Qy        780 TTTTGAAAAC 789
              ||   || ||
Db        868 TTGGAAACAC 877
```

```
RESULT 4
AX699793
LOCUS       AX699793                693 bp    DNA     linear   PAT 03-APR-2003
DEFINITION  Sequence 61 from Patent WO02070703.
ACCESSION   AX699793
VERSION     AX699793.1  GI:29500268
KEYWORDS    .
SOURCE      Porites murrayensis
  ORGANISM  Porites murrayensis
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
            Fungiina; Poritidae; Porites.
REFERENCE   1
  AUTHORS   Karan,M., Brugliera,F., Mason,J., Jones,E.L., Dove,S.G.,
            Hoegh-Guldberg,I.O. and Prescott,M.
  TITLE     Cell visual characteristic-modifying sequences
  JOURNAL   Patent: WO 02070703-A 61 12-SEP-2002;
            NUFARM AUSTRALIA LIMITED (AU) ; The University of Queensland (AU)
FEATURES             Location/Qualifiers
     source          1. .693
                     /organism="Porites murrayensis"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:227229"
     CDS             <1. .663
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD87465.1"
                     /db_xref="GI:29500269"
                     /db_xref="REMTREMBL:CAD87465"
                     /translation="SVIAKQMTYKVYMSGTVNGHYFEVQGDGKGKPYEGEQTVKLTVT
                     KGGPLPFAWDILSPQSQYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFKDGAVCTV
                     SNDSSIQGNCFIYNVKFSGLNFPPNGPVMQKKTQGWEPNTERLFARDGVLIGNNFMAL
                     KLEGGGHYLCEFKSTYKAKKPVMMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPV
                     VA"
ORIGIN

  Query Match             78.2%;  Score 657.8;  DB 6;  Length 693;
  Best Local Similarity   96.8%;  Pred. No. 1.5e-152;
  Matches  671;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db          1 TCCGTTATCGCTAAACAGATGACCTACAAGGTTTATATGTCAGGCACGGTCAATGGACAC 60

Qy         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||| |
Db         61 TACTTTGAGGTCCAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAAG 120

Qy        121 CTGGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGT 180
              ||  ||||||||||||||||||||||||||||||||||||||||||| |||||||||| ||
Db        121 CTCACTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTCTATCACCACAGAGT 180

Qy        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240

Qy        241 TCATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300
              |||||||| | | |||||||||||||||||||||||||||||  |||||||||||||||||
Db        241 TCATTCCCTGAGGGATATACATGGGAGAGGATCATGAACTTCAAAGATGGTGCAGTGTGT 300

Qy        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACAATGTCAAGTTCTCT 360

Qy        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAAAAGAAGACACAGGGCTGGGAACCC 420

Qy        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db        421 AACACTGAGCGTCTCTTTGCACGAGATGGAGTGCTGATAGGAAACAACTTTATGGCTCTG 480

Qy        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 540
              |||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        481 AAGTTGGAAGGAGGTGGTCATTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 540

Qy        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600
              ||||||| |||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        541 CCTGTGATGATGCCAGGGTATCACTATGTTGACCGCAAATTGGATGTAACCAATCACAAC 600

Qy        601 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660
              |||||||||||||||||||||||||||||| |||||||| ||||||||||||||||||||
Db        601 AAGGATTACACTTCCGTTGAGCAGTGTGAGATTTCCATCGCACGCAAACCTGTGGTCGCC 660

Qy        661 TGCCGTTTTTTCAGAGTCAAATCAAGGCACAAA 693
              || ||||||||||||||||||||||||||||||
Db        661 TGACGTTTTTTCAGAGTCAAATCAAGGCACAAA 693


RESULT 5
AX699755
LOCUS       AX699755                663 bp    DNA     linear   PAT 02-APR-2003
DEFINITION  Sequence 23 from Patent WO02070703.
ACCESSION   AX699755
VERSION     AX699755.1  GI:29500230
KEYWORDS    .
SOURCE      Acropora aspera
  ORGANISM  Acropora aspera
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
            Astrocoeniina; Acroporidae; Acropora.
REFERENCE   1
  AUTHORS   Karan,M., Brugliera,F., Mason,J., Jones,E.L., Dove,S.G.,
            Hoegh-Guldberg,I.O. and Prescott,M.
  TITLE     Cell visual characteristic-modifying sequences
  JOURNAL   Patent: WO 02070703-A 23 12-SEP-2002;
            NUFARM AUSTRALIA LIMITED (AU) ; The University of Queensland (AU)
FEATURES             Location/Qualifiers
     source          1. .663
                     /organism="Acropora aspera"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:140238"
     CDS             <1. .663
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD87446.1"
                     /db_xref="GI:29500231"
                     /db_xref="REMTREMBL:CAD87446"
                     /translation="SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLTVT
                     KGGPLPFAWDILSPQSQYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTV
                     SNDSSIQGNCFIYHVKFSGLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMAL
                     KLEGGGHYLCEFKSTYKAKKPVRMPGYHYVDRKLDVTNHNKDYTSVEQCEISIARKPV
                     VA"
ORIGIN

  Query Match             77.6%;  Score 652.4;  DB 6;  Length 663;
  Best Local Similarity   99.1%;  Pred. No. 3.3e-151;
  Matches  656;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60

Qy         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120

Qy        121 CTGGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGT 180
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGACTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTCA 180

Qy        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240

Qy        241 TCATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300
              |||||||||| | |||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 TCATTCCCGGAGGGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300

Qy        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360

Qy        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420

Qy        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480

Qy        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 540

Qy        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCTGTGAGGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600

Qy        601 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660

Qy        661 TG 662
              ||
Db        661 TG 662


RESULT 6
AX699783
LOCUS       AX699783                660 bp    DNA     linear   PAT 02-APR-2003
DEFINITION  Sequence 51 from Patent WO02070703.
ACCESSION   AX699783
VERSION     AX699783.1  GI:29500258
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1
  AUTHORS   Karan,M., Brugliera,F., Mason,J., Jones,E.L., Dove,S.G.,
            Hoegh-Guldberg,I.O. and Prescott,M.
  TITLE     Cell visual characteristic-modifying sequences
  JOURNAL   Patent: WO 02070703-A 51 12-SEP-2002;
            NUFARM AUSTRALIA LIMITED (AU) ; The University of Queensland (AU)
FEATURES             Location/Qualifiers
     source          1. .660
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
                     /note="Millepora sp. (Hydrozoan)"
     CDS             <1. .>660
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD87460.1"
                     /db_xref="GI:29500259"
                     /db_xref="REMTREMBL:CAD87460"
                     /translation="SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLTVT
                     KGGPLPFAWDILSPQSQYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTV
                     SNDSSIQGNCFIYHVKFSGLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMAL
                     KLEGGGHYLCEFKSTYKARKPVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPV
                     VA"
ORIGIN

  Query Match           77.1%;  Score 648.8;  DB 6;  Length 660;
  Best Local Similarity   98.9%;  Pred. No. 2.6e-150;
  Matches  653;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60

Qy         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120

Qy        121 CTGGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGT 180
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGACTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTCA 180

Qy        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240

Qy        241 TCATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300
              |||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCATTCCCGGAGGGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300

Qy        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360

Qy        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420

Qy        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480

Qy        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAGGAAG 540

Qy        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600

Qy        601 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        601 AAGGATTACACTTCCGTTGAGCAGCGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660


RESULT 7
AX699753
LOCUS       AX699753                660 bp    DNA     linear   PAT 02-APR-2003
DEFINITION  Sequence 21 from Patent WO02070703.
ACCESSION   AX699753
VERSION     AX699753.1  GI:29500228
KEYWORDS    .
SOURCE      Acropora aspera
  ORGANISM  Acropora aspera
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
            Astrocoeniina; Acroporidae; Acropora.
REFERENCE   1
  AUTHORS   Karan,M., Brugliera,F., Mason,J., Jones,E.L., Dove,S.G.,
            Hoegh-Guldberg,I.O. and Prescott,M.
  TITLE     Cell visual characteristic-modifying sequences
  JOURNAL   Patent: WO 02070703-A 21 12-SEP-2002;
            NUFARM AUSTRALIA LIMITED (AU) ; The University of Queensland (AU)
FEATURES             Location/Qualifiers
     source          1. .660
                     /organism="Acropora aspera"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:140238"
     CDS             <1. .>660
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD87445.1"
                     /db_xref="GI:29500229"
                     /db_xref="REMTREMBL:CAD87445"
                     /translation="SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLTVT
                     KGGPLPFAWDILSPQSQYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTV
                     SNDSSIQGNCFIYHVKFSGLNFPPNGPVMQKKTQGWEPNTERLLARDGMLIGNNFMAL
```

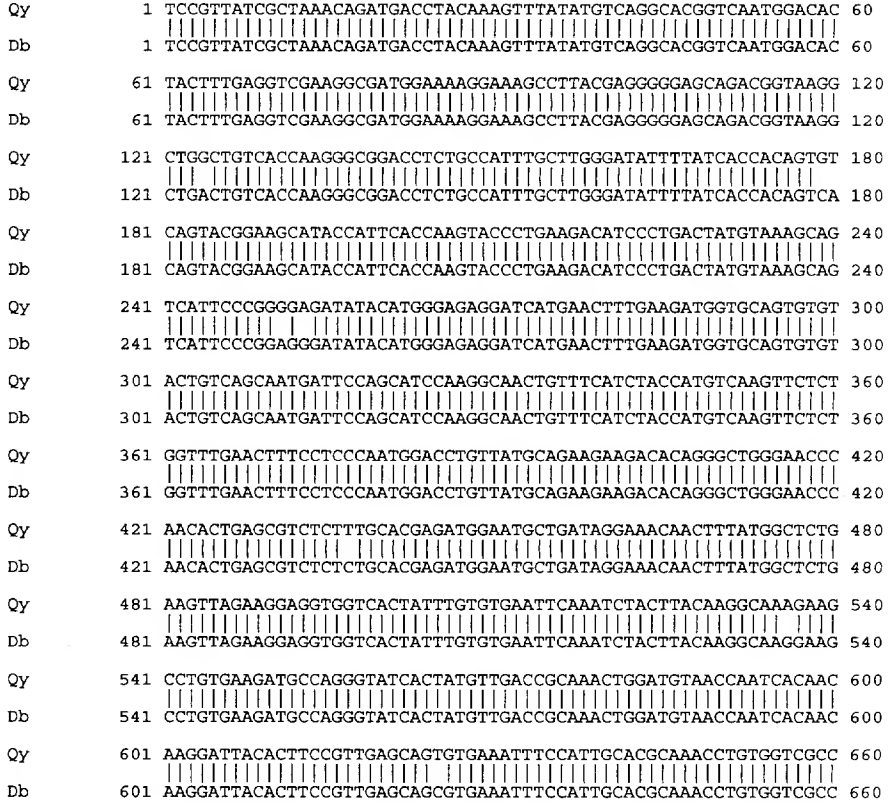
KLEGGGHYLCEFKSTYKARKPVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPV
VA"
ORIGIN

Query Match 77.0%; Score 647.2; DB 6; Length 660;
Best Local Similarity 98.8%; Pred. No. 6.5e-150;
Matches 652; Conservative 0; Mismatches 8; Indels 0; Gaps 0;



RESULT 8
AX699763
LOCUS AX699763 660 bp DNA linear PAT 02-APR-2003
DEFINITION Sequence 31 from Patent WO02070703.
ACCESSION AX699763
VERSION AX699763.1 GI:29500238
KEYWORDS .
SOURCE Caulastraea furcata
ORGANISM Caulastraea furcata
Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
Faviina; Faviidae; Caulastraea.
REFERENCE 1
AUTHORS Karan,M., Brugliera,F., Mason,J., Jones,E.L., Dove,S.G.,
Hoegh-Guldberg,I.O. and Prescott,M.
TITLE Cell visual characteristic-modifying sequences
JOURNAL Patent: WO 02070703-A 31 12-SEP-2002;
NUFARM AUSTRALIA LIMITED (AU) ; The University of Queensland (AU)
FEATURES Location/Qualifiers
source 1. .660
/organism="Caulastraea furcata"
/mol_type="unassigned DNA"
/db_xref="taxon:46696"
CDS <1. .>660
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD87450.1"
/db_xref="GI:29500239"
/db_xref="REMTREMBL:CAD87450"
/translation="SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLTVT
KGGPLPFAWDILSPQSQYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTV
SNDSSIQGNCFIYHVKFSGLNFPPNGPVMQKKTQGWEPNTERLSARDGMLIGNNFMAL
KLEGGGHYLCEFKSTYKARKPVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPV
VA"
ORIGIN

Query Match 77.0%; Score 647.2; DB 6; Length 660;
Best Local Similarity 98.8%; Pred. No. 6.5e-150;
Matches 652; Conservative 0; Mismatches 8; Indels 0; Gaps 0;



RESULT 9
AX699785
LOCUS AX699785 660 bp DNA linear PAT 02-APR-2003
DEFINITION Sequence 53 from Patent WO02070703.
ACCESSION AX699785
VERSION AX699785.1 GI:29500260
KEYWORDS .
SOURCE unidentified
ORGANISM unidentified

```
            unclassified.
REFERENCE   1
  AUTHORS   Karan,M., Brugliera,F., Mason,J., Jones,E.L., Dove,S.G.,
            Hoegh-Guldberg,I.O. and Prescott,M.
  TITLE     Cell visual characteristic-modifying sequences
  JOURNAL   Patent: WO 02070703-A 53 12-SEP-2002;
            NUFARM AUSTRALIA LIMITED (AU) ; The University of Queensland (AU)
FEATURES             Location/Qualifiers
     source          1. .660
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
                     /note="Millepora sp. (Hydrozoan)"
     CDS             <1. .>660
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD87461.1"
                     /db_xref="GI:29500261"
                     /db_xref="REMTREMBL:CAD87461"
                     /translation="SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLTVT
                     KGGPLPFAWDILSPQSQYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTV
                     SNDSSIQGNCFIYHVKFSGLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMAL
                     KLEGGGHYLCEFKSTYKARKPVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPV
                     VA"
ORIGIN

  Query Match             77.0%;  Score 647.2;  DB 6;  Length 660;
  Best Local Similarity   98.8%;  Pred. No. 6.5e-150;
  Matches  652;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60

Qy         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120

Qy        121 CTGGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGT 180
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGACTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTCA 180

Qy        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240

Qy        241 TCATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300
              |||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCATTCCCGGAGGGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300

Qy        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360

Qy        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420

Qy        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480

Qy        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 540
              |||||||||||||||| ||||||||||||||||||||||||||||||||||| ||||
Db        481 AAGTTAGAAGGAGGTGGCCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAGGAAG 540

Qy        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600

Qy        601 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        601 AAGGATTACACTTCCGTTGAGCAGCGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660
```

```
RESULT 10
AX699923
LOCUS       AX699923                669 bp    DNA     linear   PAT 02-APR-2003
DEFINITION  Sequence 191 from Patent WO02070703.
ACCESSION   AX699923
VERSION     AX699923.1  GI:29500381
KEYWORDS    .
SOURCE      Discosoma sp.
  ORGANISM  Discosoma sp.
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
            Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE   1
  AUTHORS   Karan,M., Brugliera,F., Mason,J., Jones,E.L., Dove,S.G.,
            Hoegh-Guldberg,I.O. and Prescott,M.
  TITLE     Cell visual characteristic-modifying sequences
  JOURNAL   Patent: WO 02070703-A 191 12-SEP-2002;
            NUFARM AUSTRALIA LIMITED (AU) ; The University of Queensland (AU)
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Discosoma sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:86600"
ORIGIN

  Query Match             77.0%;  Score 647.2;  DB 6;  Length 669;
  Best Local Similarity   98.8%;  Pred. No. 6.5e-150;
  Matches  652;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 63

Qy         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 123

Qy        121 CTGGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGT 180
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 CTGACTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTCA 183

Qy        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 243

Qy        241 TCATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300
              |||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db        244 TCATTCCCGGAGGGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 303

Qy        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 363

Qy        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 423

Qy        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        424 AACACTGAGCGTCTCTTAGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 483

Qy        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        484 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAGGAAG 543

Qy        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 603

Qy        601 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        604 AAGGATTACACTTCCGTTGAGCAGCGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 663
```

```
RESULT 11
AX699759
LOCUS       AX699759                660 bp    DNA     linear   PAT 02-APR-2003
DEFINITION  Sequence 27 from Patent WO02070703.
ACCESSION   AX699759
VERSION     AX699759.1  GI:29500234
KEYWORDS    .
SOURCE      Acanthastrea echinata
  ORGANISM  Acanthastrea echinata
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
            Faviina; Mussidae; Acanthastrea.
REFERENCE   1
  AUTHORS   Karan,M., Brugliera,F., Mason,J., Jones,E.L., Dove,S.G.,
            Hoegh-Guldberg,I.O. and Prescott,M.
  TITLE     Cell visual characteristic-modifying sequences
  JOURNAL   Patent: WO 02070703-A 27 12-SEP-2002;
            NUFARM AUSTRALIA LIMITED (AU) ; The University of Queensland (AU)
FEATURES             Location/Qualifiers
     source          1. .660
                     /organism="Acanthastrea echinata"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:51051"
     CDS             <1. .>660
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD87448.1"
                     /db_xref="GI:29500235"
                     /db_xref="REMTREMBL:CAD87448"
                     /translation="SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVRLTVT
                     KGGPLPFAWDILSPQSQYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTV
                     SNDSSIQGNCFIYHVKFSGLNFPPNGPVMQKKTQGWEPNTERLSARDGMLIGNNFMAL
                     KLEGGGHYLCEFKSTYKARKPVKMPGYHCVDRKLDVTNHNKDYTSVEQREISIARKPV
                     VA"
ORIGIN

  Query Match           76.8%;  Score 645.6;  DB 6;  Length 660;
  Best Local Similarity  98.6%;  Pred. No. 1.6e-149;
  Matches  651;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60

Qy         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120

Qy        121 CTGGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGT 180
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGACTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTCA 180

Qy        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240

Qy        241 TCATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300
              |||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCATTCCCGGAGGGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300

Qy        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360

Qy        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420

Qy        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACACTGAGCGTCTCTCTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480

Qy        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAGGAAG 540

Qy        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        541 CCTGTGAAGATGCCAGGGTATCACTGTGTTGACCGCAAACTGGATGTAACCAATCACAAC 600

Qy        601 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        601 AAGGATTACACTTCCGTTGAGCAGCGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660


RESULT 12
AX699921
LOCUS       AX699921                669 bp    DNA     linear   PAT 02-APR-2003
DEFINITION  Sequence 189 from Patent WO02070703.
ACCESSION   AX699921
VERSION     AX699921.1  GI:29500380
KEYWORDS    .
SOURCE      Acropora aspera
  ORGANISM  Acropora aspera
            Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
            Astrocoeniina; Acroporidae; Acropora.
REFERENCE   1
  AUTHORS   Karan,M., Brugliera,F., Mason,J., Jones,E.L., Dove,S.G.,
            Hoegh-Guldberg,I.O. and Prescott,M.
  TITLE     Cell visual characteristic-modifying sequences
  JOURNAL   Patent: WO 02070703-A 189 12-SEP-2002;
            NUFARM AUSTRALIA LIMITED (AU) ; The University of Queensland (AU)
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Acropora aspera"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:140238"
ORIGIN

  Query Match           76.4%;  Score 642.4;  DB 6;  Length 669;
  Best Local Similarity  98.3%;  Pred. No. 1e-148;
  Matches  649;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db          4 TCCGTTATCGCTAAACAGATGACCTACAAGGTTTATATGTCAGGCACGGTCAATGGACAC 63

Qy         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db         64 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAAG 123

Qy        121 CTGGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGT 180
              ||  |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        124 CTCACTGTCACCAAGGGTGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTCA 183

Qy        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        184 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCGGACTATGTAAAGCAG 243

Qy        241 TCATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300
              |||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db        244 TCATTCCCGGAGGGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 303

Qy        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        304 ACTGTCAGCAATGACTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 363

Qy        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 423

Qy        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 483
```

```
Qy        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 543

Qy        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 603

Qy        601 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 663


RESULT 13
AX699927
LOCUS       AX699927                669 bp    DNA     linear   PAT 02-APR-2003
DEFINITION  Sequence 195 from Patent WO02070703.
ACCESSION   AX699927
VERSION     AX699927.1  GI:29500383
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1
  AUTHORS   Karan,M., Brugliera,F., Mason,J., Jones,E.L., Dove,S.G.,
            Hoegh-Guldberg,I.O. and Prescott,M.
  TITLE     Cell visual characteristic-modifying sequences
  JOURNAL   Patent: WO 02070703-A 195 12-SEP-2002;
            NUFARM AUSTRALIA LIMITED (AU) ; The University of Queensland (AU)
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
                     /note="Tubastrea sp."
ORIGIN

  Query Match             76.2%;  Score 640.8;  DB 6;  Length 669;
  Best Local Similarity   98.2%;  Pred. No. 2.5e-148;
  Matches  648;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          4 TCCGTTATCGCTAAACAGATGACCTACAAGGTTTATATGTCAGGCACGGTCAATGGACAC 63

Qy         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db         64 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAAG 123

Qy        121 CTGGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGT 180
              ||  |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        124 CTCACTGTCACCAAGGGTGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTCA 183

Qy        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        184 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCGGACTATGTAAAGCAG 243

Qy        241 TCATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300
              |||||||||| | |||||||||||||||||||||||||||||||||||||||||||||||
Db        244 TCATTCCCGGAGGGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 303

Qy        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        304 ACTGTCAGCAATGACTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 363

Qy        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 423

Qy        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 483
```

```
Qy        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 543

Qy        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 603

Qy        601 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCGCGCAAACCTGTGGTCGCC 663


RESULT 14
AX699933
LOCUS       AX699933                669 bp    DNA     linear   PAT 02-APR-2003
DEFINITION  Sequence 201 from Patent WO02070703.
ACCESSION   AX699933
VERSION     AX699933.1  GI:29500386
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1
  AUTHORS   Karan,M., Brugliera,F., Mason,J., Jones,E.L., Dove,S.G.,
            Hoegh-Guldberg,I.O. and Prescott,M.
  TITLE     Cell visual characteristic-modifying sequences
  JOURNAL   Patent: WO 02070703-A 201 12-SEP-2002;
            NUFARM AUSTRALIA LIMITED (AU) ; The University of Queensland (AU)
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
                     /note="Tubastrea sp."
ORIGIN

  Query Match             76.2%;  Score 640.8;  DB 6;  Length 669;
  Best Local Similarity   98.2%;  Pred. No. 2.5e-148;
  Matches  648;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          4 TCCGTTATCGCTAAACAGATGACCTACAAGGTTTATATGTCAGGCACGGTCAATGGACAC 63

Qy         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 123

Qy        121 CTGGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 CTGGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGT 183

Qy        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        184 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCGG 243

Qy        241 TCATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300
              |||||||||| | ||| |||||||||||||||||||||||||||||||||||||||||||||
Db        244 TCATTCCCGGAGGGATTTACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 303

Qy        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 363

Qy        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 423

Qy        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480
              || ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        424 CACTCTGAGCGTCTCTTTGCACGAGACGGAATGCTGATAGGAAACAACTTTATGGCTCTG 483
```

```
Qy        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 540
              ||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||||
Db        484 AAGTTAGAAGGAGGCGGTCACTATTTGTGTGAATTCAAAACTACTTACAAGGCAAAGAAG 543

Qy        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600
              |||||||||||||||||||||| |||||||||||||||||||||||||| ||||||||||
Db        544 CCTGTGAAGATGCCAGGGTATCATTATGTTGACCGCAAACTGGATGTAATCAATCACAAC 603

Qy        601 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 663


RESULT 15
AX699781
LOCUS       AX699781                660 bp    DNA     linear   PAT 02-APR-2003
DEFINITION  Sequence 49 from Patent WO02070703.
ACCESSION   AX699781
VERSION     AX699781.1  GI:29500256
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1
  AUTHORS   Karan,M., Brugliera,F., Mason,J., Jones,E.L., Dove,S.G.,
            Hoegh-Guldberg,I.O. and Prescott,M.
  TITLE     Cell visual characteristic-modifying sequences
  JOURNAL   Patent: WO 02070703-A 49 12-SEP-2002;
            NUFARM AUSTRALIA LIMITED (AU) ; The University of Queensland (AU)
FEATURES             Location/Qualifiers
     source          1. .660
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
                     /note="Millepora sp. (Hydrozoan)"
     CDS             <1. .>660
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD87459.1"
                     /db_xref="GI:29500257"
                     /db_xref="REMTREMBL:CAD87459"
                     /translation="SVIAKQMTYKVYMSGTVNGHYFEVEGDGKGKPYEGEQTVKLTVT
                     KGGPLPFAWDILSPQSQYGSIPFTKYPEDIPDYVKQSFPEGYTWERIMNFEDGAVCTV
                     SNDSSIQGNCFIYHVKFSGLNFPPNGPVMQKKTQGWEPNTERLFARDGMLIGNNFMAL
                     KLEGGGHYLCEFKSTYKARKPVKMPGYHYVDRKLDVTNHNKDYTSVEQREISIARKPV
                     VA"
ORIGIN

  Query Match           76.0%;  Score 639.2;  DB 6;  Length 660;
  Best Local Similarity   98.0%;  Pred. No. 6.3e-148;
  Matches  647;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 TCCGTTATCGCTAAACAGATGACCTACAAAGTTTATATGTCAGGCACGGTCAATGGACAC 60
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db          1 TCCGTTATCGCTAAACAGATGACCTACAAGGTTTATATGTCAGGCACGGTCAATGGACAC 60

Qy         61 TACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAGG 120
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||| |
Db         61 TACTTTGAGGTTGAAGGCGATGGAAAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAAG 120

Qy        121 CTGGCTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTTTATCACCACAGTGT 180
              ||  |||||||||||||||||||||||||||||||||||||||| |||||||||| ||
Db        121 CTCACTGTCACCAAGGGCGGACCTCTGCCATTTGCTTGGGATATTCTATCACCACAGAGT 180

Qy        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAG 240

Qy        241 TCATTCCCGGGGAGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300
              |||||||| | | ||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCATTCCCTGAGGGATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT 300

Qy        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACTGTCAGCAATGATTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCT 360

Qy        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGTTTGAACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCC 420

Qy        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACACTGAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTG 480

Qy        481 AAGTTAGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAAG 540
              |||||||||||||||||||||||||| |||||||||||||||||||||||||| ||||
Db        481 AAGTTAGAAGGAGGTGGTCACTATTTATGTGAATTCAAATCTACTTACAAGGCAAGGAAG 540

Qy        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATCACAAC 600

Qy        601 AAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        601 AAGGATTACACTTCCGTTGAGCAGCGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCC 660
```

Search completed: August 13, 2004, 19:12:33
Job time : 3676 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 13, 2004, 11:19:50 ; Search time 2783.5 Seconds
(without alignments)
9022.487 Million cell updates/sec

Title: US-09-890-463-6
Perfect score: 841
Sequence: 1 tccgttatcgctaaacagat.....aaaagcggccgctcgaatta 841

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	57	6.8	866	14	CK159216	CK159216 FGAS04062
	2	56	6.7	297	14	CF805258	CF805258 lad65f03.
	3	56	6.7	820	13	BU563278	BU563278 AGENCOURT
c	4	55.6	6.6	667	29	CNS02YZV	AL220324 Tetraodon
	5	55	6.5	359	29	CNS007CK	AL066995 Drosophil
	6	55	6.5	683	12	BI527017	BI527017 602927295
c	7	55	6.5	1201	13	BX398622	BX398622 BX398622
	8	54.8	6.5	635	14	CD771722	CD771722 AGENCOURT
	9	54.6	6.5	362	14	CB958074	CB958074 AGENCOURT
	10	54.6	6.5	752	13	BU567098	BU567098 AGENCOURT
	11	54.4	6.5	278	13	BQ566832	BQ566832 gi70h11.y
c	12	54.4	6.5	539	13	BX403488	BX403488 BX403488
c	13	54.2	6.4	872	14	CK155159	CK155159 FGAS03388
	14	54.2	6.4	918	9	AL558886	AL558886 AL558886
	15	54.2	6.4	1101	29	CNS00KB5	AL077453 Drosophil
c	16	53.8	6.4	242	9	AI796743	AI796743 wa14d04.x
	17	53.8	6.4	314	10	BE047863	BE047863 tz43c10.y
	18	53.8	6.4	330	9	AU033588	AU033588 AU033588
	19	53.8	6.4	418	14	CF309143	CF309143 ABF--03-D
c	20	53.8	6.4	722	9	AL663605	AL663605 AL663605
c	21	53.8	6.4	794	14	CK197676	CK197676 FGAS00615
	22	53.6	6.4	399	12	BI746239	BI746239 rm24f05.y
c	23	53.6	6.4	855	14	CK196558	CK196558 FGAS00501
	24	53.4	6.3	249	14	CK240710	CK240710 rx12e04.y
	25	53.4	6.3	528	14	CA324119	CA324119 UI-M-FY0-
	26	53.4	6.3	901	13	BU555445	BU555445 AGENCOURT
c	27	53.2	6.3	852	14	CK194319	CK194319 FGAS00274
	28	53	6.3	278	14	CF423560	CF423560 lac96e07.
c	29	53	6.3	1201	13	BX403935	BX403935 BX403935
	30	52.8	6.3	278	12	BG609931	BG609931 324335 MA
	31	52.8	6.3	417	10	BF294370	BF294370 004PbD05
	32	52.8	6.3	422	9	AU262401	AU262401 AU262401
	33	52.8	6.3	441	12	BG662904	BG662904 DRA05B01
c	34	52.8	6.3	561	14	CF328004	CF328004 NACL--02-
	35	52.8	6.3	675	14	CD641892	CD641892 AGENCOURT
c	36	52.8	6.3	867	29	CNS0054A	AL057618 Drosophil
c	37	52.8	6.3	873	14	CK195879	CK195879 FGAS00432
	38	52.8	6.3	888	13	BU946204	BU946204 AGENCOURT
c	39	52.6	6.3	851	14	CK151995	CK151995 FGAS03482
c	40	52.6	6.3	937	13	BX328575	BX328575 BX328575
c	41	52.6	6.3	1184	13	BX446507	BX446507 BX446507
	42	52.4	6.2	223	10	AW433218	AW433218 sh54a03.y
	43	52.4	6.2	521	12	BI378622	BI378622 BFLG1_000
	44	52.4	6.2	812	12	BM985610	BM985610 2_B07_T3
c	45	52.4	6.2	868	14	CK152086	CK152086 FGAS03499

ALIGNMENTS

RESULT 1
CK159216/c
LOCUS CK159216 866 bp mRNA linear EST 05-DEC-2003
DEFINITION FGAS040623 Triticum aestivum FGAS: TaLt5 Triticum aestivum cDNA, mRNA sequence.
ACCESSION CK159216
VERSION CK159216.1 GI:38985155
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 866)
AUTHORS Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D., Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A., Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D., Penniket,C., Roach,J.L. and Sarhan,F.
TITLE Functional Genomics of Abiotic Stress In Wheat and Canola Crops
JOURNAL Unpublished (2003)
COMMENT Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033

```
            Email: fgas_ests@cs.usask.ca
            This sequence is the direct result of the Base calling software
            Phred (default parameters). It is the raw base calls. To aid in the
            identification of the high quality insert the software Lucy
            (default parameters) has been run on this sequence. Lucy identified
            the region [124,600].
            Plate: TaLt538  row: B  column: 11.
FEATURES              Location/Qualifiers
     source           1. .866
                      /organism="Triticum aestivum"
                      /mol_type="mRNA"
                      /cultivar="Wheat line PI 178383"
                      /db_xref="taxon:4565"
                      /lab_host="DH5 alpha"
                      /clone_lib="Triticum aestivum FGAS: TaLt5"
                      /note="Organ: Crown; Vector: pGEM-T; SSH (suppression
                      subtractive hybridization) cDNA library from genotype
                      PI178383 cold hardened at 2 C for 21 days and 49 days
                      (equal amount of cDNA pooled together before subtraction,
                      tester) and subtracted against genotype Norstar cold
                      hardened at 2 C for 1 day (24 H)(driver). Modified Smart
                      cDNA (Clontech)priming and non-directional cloning"
ORIGIN

  Query Match             6.8%;  Score 57;  DB 14;  Length 866;
  Best Local Similarity   60.8%;  Pred. No. 7.5;
  Matches   93;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy        673 AGAGTCAAATCAAGGCACAAATACGCAGTGGCGTAAAAAACGTAGATTCTGATTTTAGCT 732
              | | | ||| |||| | || | |   |   | |||||| | | || | |||||
Db        384 AAAATAAAAACAAGAAAAAATTTTGGTTTTAAGCAAAAAAAGGGGTTTTTAATTTTTGTA 325

Qy        733 TATAGAAGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAAACTTTTGAAAACAAC 792
              | | || |  |||  |||    ||| ||  | ||  |  || ||||  || ||
Db        324 AAAAAAAAAAAAAACTGGGAAAATAAAAAAAAATTAAAACACGGAATTTTTTTTAAAAAA 265

Qy        793 GCCATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 825
                | ||||||||||||||||||||||||||||||
Db        264 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 232


RESULT 2
CF805258
LOCUS       CF805258                297 bp    mRNA    linear   EST 24-OCT-2003
DEFINITION  lad65f03.y1 Gastric Epithelial Progenitor 2 Mus musculus cDNA 5',
            mRNA sequence.
ACCESSION   CF805258
VERSION     CF805258.1  GI:37974718
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 297)
  AUTHORS   Tidwell,R., Clifton,S., Marra,M., Hillier,L., Pape,D., Martin,J.,
            Wylie,T., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennet,J.,
            Ronko,I., Tsagareishvili,R., Belaygorod,L., Grow,A., Maguire,L.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Stem cell EST Project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Jeff Gordon and Mike Lovett
            WashU, Human Genetics Division
            Washington University School of Medicine
            1st strand of cDNA was synthesized with reverse transcriptase and
            oligo(dT) beads, then cDNA was amplified by PCR using modified
            SMART primers. The final cDNA was cloned in pAMP1 vector in
            annealing reaction with Uracil DNA Glycosylase (UDG). Library
            constructed by Y.Korshunova and M. Lovett. Library materials
            provided by Mills JC & Gordon JI.
            Putative full length read
            vector to vector length is
            Seq primer: -40RP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .297
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10090"
                      /tissue_type="Laser-captured isthmal cells from tox176
                      transgenic mice"
                      /dev_stage="adult"
                      /lab_host="DH5alpha"
                      /clone_lib="Gastric Epithelial Progenitor 2"
                      /note="Vector: pAMP1; 1st strand of cDNA was synthesized
                      with reverse transcriptase and oligo(dT) beads, then cDNA
                      was amplified by PCR using modified SMART primers. The
                      final cDNA was cloned in pAMP1 vector in annealing
                      reaction with Uracil DNA Glycosylase (UDG). Library
                      constructed by Y.Korshunova and M. Lovett. Library
                      materials provided by Mills JC & Gordon JI."
ORIGIN

  Query Match             6.7%;  Score 56;  DB 14;  Length 297;
  Best Local Similarity   58.3%;  Pred. No. 18;
  Matches   98;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        658 GCCTGCCGTTTTTTCAGAGTCAAATCAAGGCACAAATACGCAGTGGCGTAAAAAACGTAG 717
              |||| || |     ||||| | |      ||  || |||  ||| ||   ||| |  ||
Db        128 GCCTTCCCTCCACACAGAGACCAGAGTGGGGTCACATAAACAGAGGGAAAAAGAGAGTCT 187

Qy        718 ATTCTGATTTTAGCTTATAGAAGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAA 777
              ||| |  | | |   || |||| |   |  || ||    ||| ||  |  ||  |  ||
Db        188 ATTTTTGTGTAATAATAAAGAATTTCTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 247

Qy        778 ACTTTTGAAAACAACGCCATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 825
              |      |||| ||    | ||||||||||||||||||||||||||||
Db        248 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 295


RESULT 3
BU563278
LOCUS       BU563278                820 bp    mRNA    linear   EST 16-SEP-2002
DEFINITION  AGENCOURT_10248984 NIH_MGC_143 Mus musculus cDNA clone
            IMAGE:6596463 5', mRNA sequence.
ACCESSION   BU563278
VERSION     BU563278.1  GI:22913578
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 820)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
             cDNA Library Preparation: Michael Brownstein Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2826  row: b  column: 15
            High quality sequence stop: 312.
FEATURES              Location/Qualifiers
     source           1. .820
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:6596463"
                      /lab_host="DH10B (T1-phage-resistant)"
                      /clone_lib="NIH_MGC_143"
                      /note="Organ: Brain; Vector: pDNR-LIB; Site_1: SfiI
                      (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); cDNA made
                      by oligo-dT priming and directionally cloned. 5' and 3'
```

adaptors were used in cloning as follows: 5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.2-0.5 kb size fraction (other fractions present in NIH_MGC_144). Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 6.7%; Score 56; DB 13; Length 820;
Best Local Similarity 53.7%; Pred. No. 11;
Matches 116; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```
Qy        610 ACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCCTGCCGTTTT 669
              ||| |  ||  | ||      | ||| ||| |  |||| |   |||| ||   |     |
Db         39 ACTGCTCTTCTGAAGGTCCTGAGTTCAATTCCCAGCAACCACATGGTGGCTCACAACCAT 98

Qy        670 TTCAGAGTCAAATCAAGGCACAAATACGCAGTGGCGTAAAAAACGTAGATTCTGATTTTA 729
              ||   |    |    | || |   | |||| |  || | |  || | | | ||  ||
Db         99 TTGTAATGAGATCTGATGCCCTCTTCTGGAGTGTCTGAAGACAGCTACAGTGTAATCCTA 158

Qy        730 GCTTATAGAAGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAAACTTTTGAAAAC 789
                || | || ||       || ||    ||| ||  |  ||  |  |||       ||||
Db        159 TAATAAATAATTAAATCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 218

Qy        790 AACGCCATAAAAAAAAAAAAAAAAAAAAAAAAAAA 825
              ||    | ||||||||||||||||||||||||||||
Db        219 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 254
```

RESULT 4
CNS02YZV/c
LOCUS CNS02YZV 667 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 181N15 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL220324
VERSION AL220324.1 GI:7879143
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE 1
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
PUBMED 10899143
REFERENCE 3 (bases 1 to 667)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .667
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="181N15"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG181CG08SP1~end : PUC-Ori"

ORIGIN

Query Match 6.6%; Score 55.6; DB 29; Length 667;
Best Local Similarity 48.1%; Pred. No. 14;
Matches 76; Conservative 24; Mismatches 58; Indels 0; Gaps 0;

```
Qy        678 CAAATCAAGGCACAAATACGCAGTGGCGTAAAAAACGTAGATTCTGATTTTAGCTTATAG 737
              |:||:  | : : |||:|  :::|::   |||||:   | |  : |:|:  :   | |
Db        201 CWAAWACAARARAAAAWAAWMWRTRRAAAAAAAAAMAAAAAAAAMMGRTWRAWMGAAAAAA 142

Qy        738 AAGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAAACTTTTGAAAACAACGCCAT 797
              ||  | ::|: |  |:    :||||||  |  ||| |  |||       |:|: ||    |
Db        141 AAMAAARMAMMACMAMMAAARAACAACAAMAAATCAACAAAACAAAAAMAMAAAAAAAAA 82

Qy        798 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGGCCGCTC 835
              |||||||||||||||||||||||||||||:| :   | |||
Db         81 AAAAAAAAAAAAAAAAAAAAAAAAAAAAMAAMCAGCTCTC 44
```

RESULT 5
CNS007CK
LOCUS CNS007CK 359 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR15J06 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL066995
VERSION AL066995.1 GI:4945559
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 359)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .359
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR15J06"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

```
  Query Match             6.5%;  Score 55;  DB 29;  Length 359;
  Best Local Similarity   62.2%;  Pred. No. 24;
  Matches   74;  Conservative    7;  Mismatches   38;  Indels    0;  Gaps    0;

Qy        707 AAAAAACGTAGATTCTGATTTTAGCTTATAGAAGTAGGAACGAAGAAGTGTAAACAACCA 766
              |||:||  || |:: : || |    |||| || :|  ||: || || | || |||| ||
Db        141 AAAWAAAATATAWHAWAATATATATATATATAAAWAAAAAMAAAAAAATATATACAAACA 200

Qy        767 TTAATGATTAAACTTTTGAAAACAACGCCATAAAAAAAAAAAAAAAAAAAAAAAAAAAA 825
              | ||  | |||      |||| ||    | |||||:||||||||||||||||||||
Db        201 TAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAMAAAAAAAAAAAAAAAAAAAA 259


RESULT 6
BI527017
LOCUS       BI527017                683 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  602927295F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5059775 5',
            mRNA sequence.
ACCESSION   BI527017
VERSION     BI527017.1  GI:15351809
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11161  row: m  column: 24
            High quality sequence start: 2
            High quality sequence stop: 437.
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N-3"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5059775"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam2"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
ORIGIN

  Query Match             6.5%;  Score 55;  DB 12;  Length 683;
  Best Local Similarity   51.9%;  Pred. No. 18;
  Matches  124;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy        586 GTAACCAATCACAACAAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGC 645
              ||  | | | |  |  ||  | || | ||  || ||||||     || |  ||||| |
Db        127 GTGGCTATTAATTATTCGGTCTGCATTCCCAGTGGGCAGTGATGGCATTACTCTGCACTC 186

Qy        646 AAACCTGTGGTCGCCTGCCGTTTTTTCAGAGTCAAATCAAGGCACAAATACGCAGTGGCG 705
              | ||  | | | |       |||  | | |  ||        | |   | |     |
Db        187 TAGCCACTTGCCCCAATTTAAGTTTAGAAATTACAAGTTTCAATAATAGCTGAACCTCTG 246

Qy        706 TAAAAAACGTAGATTCTGATTTTAGCTTATAGAAGTAGGAACGAAGAAGTGTAAACAACC 765
              | ||||| ||  ||   | ||||     || | |  |  ||  || ||   | || ||
```

```
Db        247 TTAAAAATGTTAATAAAGGTTTTGTTGCATGGTAAGACAAAAAAAAAAAAAATTAAAAAAA 306

Qy        766 ATTAATGATTAAACTTTTGAAAACAACGCCATAAAAAAAAAAAAAAAAAAAAAAAAAAAA 824
              |  ||  |  |||      |||| ||   || ||||||||| ||||||||||||||||
Db        307 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAAAAAAAACAAAAAAAAAAAAAAA 365


RESULT 7
BX398622/c
LOCUS       BX398622                1201 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX398622 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI061YJ01 5-PRIME, mRNA sequence.
ACCESSION   BX398622
VERSION     BX398622.1  GI:30617636
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI061CE01QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI061YJ01"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             6.5%;  Score 55;  DB 13;  Length 1201;
  Best Local Similarity   43.5%;  Pred. No. 13;
  Matches   64;  Conservative   34;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        679 AAATCAAGGCACAAATACGCAGTGGCGTAAAAAACGTAGATTCTGATTTTAGCTTATAGA 738
              |||  ||  || ||| |           |:|||:  || ::| |  |:::: ::::| :
Db        449 AAAAAAACCCARAAAAAAAAGAAAAAAAARAAAMAATATWWTTTTTTWWWWTTWWWWATW 390

Qy        739 AGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAAACTTTTGAAAACAACGCCATA 798
              | ::  ||  :: ::   ::|| ||  | |: : :::::: :::: |||| |:    | |
Db        389 AWWWWAAAAAWWTWWWAWWWAAWAAAAAATWTWWWWWWWWWWWWWWAAAAAAWAAAAAA 330

Qy        799 AAAAAAAAAAAAAAAAAAAAAAAAAA 825
              ||||||||||||||||||||||||||
Db        329 AAAAAAAAAAAAAAAAAAAAAAAAAA 303


RESULT 8
CD771722
LOCUS       CD771722                635 bp    mRNA    linear   EST 02-JUL-2003
DEFINITION  AGENCOURT_14738271 NIH_MGC_190 Mus musculus cDNA clone
            IMAGE:30501368 5', mRNA sequence.
ACCESSION   CD771722
VERSION     CD771722.1  GI:32430224
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Yoshihiko Yamada, Takashi Nakamura, NIDCR
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM223  row: o  column: 09
            High quality sequence stop: 305.
FEATURES              Location/Qualifiers
     source           1. .635
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:30501368"
                      /lab_host="DH10B (phage-resistant)"
                      /clone_lib="NIH_MGC_190"
                      /note="Organ: Pooled - Molar; Vector: pDNR-LIB; Site_1:
                      SfiI (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc);
                      Non-normalized full-length enriched library 5' and 3'
                      adaptors were used in cloning as follows: 5' adaptor
                      sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                      5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                      C, or G and N = A, C, G, or T). Average insert size 1.71
                      kb (range 0.5-3.0 kb). 15/15 colonies contained inserts by
                      PCR. This library was enriched for full-length clones and
                      was constructed by Clontech Laboratories (Palo Alto, CA)
                      Corp."
ORIGIN

  Query Match             6.5%;  Score 54.8;  DB 14;  Length 635;
  Best Local Similarity   53.2%;  Pred. No. 20;
  Matches  116;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy        610 ACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCCTGCCGTTTT 669
              ||| |  ||  | ||    | | ||| | | |  |||| |   |||| ||   |     |
Db         41 ACTGCTCTTCCGAAGGTCCAGAGTTCAAATCCCAGCAACCACATGGTGGCTCACAACCAT 100

Qy        670 TTCAGAGTCAAATCAAGGCACAAATACGCAGTGGCGTAAAAAACGTAGATTCTGATTTTA 729
              |  |  ||    | || |   |  | |||| |  || | |  || | | |  || |
Db        101 CCCTAATGAAATCTGATGCCCTCTTCTGGAGTGTCTGAAGACAGCTACAGTGTACTTACA 160

Qy        730 GCTTATAGAAGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAAACTTTTGAAAAC 789
              | ||  |   |  ||  || ||    ||| ||  |  ||  |  |||      ||||
Db        161 TATAATCAATAAATAAATCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 220

Qy        790 AACGCCATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGC 827
              ||    | |||||||||||||||||||||||||||| |
Db        221 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAC 258


RESULT 9
CB958074
LOCUS       CB958074                362 bp    mRNA    linear   EST 29-APR-2003
DEFINITION  AGENCOURT_13785450 NIH_MGC_184 Homo sapiens cDNA clone
            IMAGE:30351140 5', mRNA sequence.
ACCESSION   CB958074
VERSION     CB958074.1  GI:30214191
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 362)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM148  row: c  column: 21
            High quality sequence stop: 337.
FEATURES              Location/Qualifiers
     source           1. .362
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:30351140"
                      /lab_host="DH10B (T1 phage-resistant)"
                      /clone_lib="NIH_MGC_184"
                      /note="Organ: Pooled-Glandular; Vector: pDNR-LIB; Site_1:
                      SfiI (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc);
                      Library is oligo-dT primed and directionally cloned. cDNA
                      was prepared from a glandular pool of tissues from thyoid,
                      parathyroid, adrenal, cortex and pineal gland.  5' and 3'
                      adaptors were used in cloning as follows: 5' adaptor
                      sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                      5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                      C, or G and N = A, C, G, or T). Average insert size 1.38
                      kb (range 0.60-3.5 kb). 15/15 colonies contained inserts
                      by PCR. This library was enriched for full-length clones
                      and was constructed by Clontech Laboratories (Palo Alto,
                      CA). Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             6.5%;  Score 54.6;  DB 14;  Length 362;
  Best Local Similarity   59.2%;  Pred. No. 28;
  Matches   93;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        669 TTTCAGAGTCAAATCAAGGCACAAATACGCAGTGGCGTAAAAAACGTAGATTCTGATTTT 728
              ||| | | ||||  ||  | |     |   | || | | ||| || ||  ||||  |||||
Db        155 TTTAAAATTCAATGCAGAGAAGTTGTTGACTGTAGGGGAAATAAAGTTAATTCAAATTTT 214

Qy        729 AGCTTATAGAAGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAAACTTTTGAAAA 788
              | | ||  |  ||  || ||    ||| ||  |  ||  |  |||      ||||
Db        215 GAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 274

Qy        789 CAACGCCATAAAAAAAAAAAAAAAAAAAAAAAAAAA 825
              ||    | |||||||||||||||||||||||||||
Db        275 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 311


RESULT 10
BU567098
LOCUS       BU567098                752 bp    mRNA    linear   EST 16-SEP-2002
DEFINITION  AGENCOURT_10393772 NIH_MGC_141 Homo sapiens cDNA clone
            IMAGE:6606668 5', mRNA sequence.
ACCESSION   BU567098
VERSION     BU567098.1  GI:22917398
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 752)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

Tissue Procurement: NCI
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLCM2852 row: k column: 20
High quality sequence stop: 400.

FEATURES Location/Qualifiers
source 1. .752
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6606668"
/tissue_type="mixed (pool of 40 RNAs)"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH_MGC_141"
/note="Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); Double-stranded cDNA was prepared from a pool of 40 cell line polyA+ RNAs (bladder - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon - 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%, kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were used in cloning as follows: 5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.2-0.5 kb size fraction (other fractions present in NIH_MGC_142). Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 6.5%; Score 54.6; DB 13; Length 752;
Best Local Similarity 59.2%; Pred. No. 19;
Matches 93; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy        669 TTTCAGAGTCAAATCAAGGCACAAATACGCAGTGGCGTAAAAAACGTAGATTCTGATTTT 728
              ||| | | ||||  ||  | |     |  | || | | ||| || ||  ||||  |||||
Db        221 TTTAAAATTCAATGCAGAGAAGTTGTTGACTGTAGGGGAAATAAAGTTAATTCAAATTTT 280

Qy        729 AGCTTATAGAAGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAAACTTTTGAAAA 788
              | | ||  |  ||  || ||    ||| ||  |  ||  |  |||       ||||
Db        281 GAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 340

Qy        789 CAACGCCATAAAAAAAAAAAAAAAAAAAAAAAAAAA 825
                ||    | ||||||||||||||||||||||||||||
Db        341 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 377
```

RESULT 11
BQ566832
LOCUS BQ566832 278 bp mRNA linear EST 19-JUN-2002
DEFINITION gi70h11.y1 Mouse Organ of Corti cDNA pBluescript Mus musculus cDNA clone gi70h11 5', mRNA sequence.
ACCESSION BQ566832
VERSION BQ566832.1 GI:21470149
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 278)
AUTHORS Kachar,B.
TITLE EST analysis of gene expression in the mouse Organ of Corti at the onset of hearing
JOURNAL Unpublished (2002)
COMMENT Contact: Kachar,B.
Structural Cell Biology
National Institute of Deafness and other Communication Disorders
50/4249 South Drive, NIH, Bethesda, MD 20892-8027, USA
Tel: 301-402-1599
Fax: 301-402-1765
Email: kacharb@nidcd.nih.gov
Plate: 70 row: h column: 11
Seq primer: M13RP1 reverse primer (ABI).

FEATURES Location/Qualifiers
source 1. .278
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="gi70h11"
/sex="male and female"
/dev_stage="Post natal day 5 to 13"
/clone_lib="Mouse Organ of Corti cDNA pBluescript"
/note="Organ: Organ of Corti; Vector: pBluescript; The organ of Corti (OC) was fine dissected from a total of 386 OC as follows: 102 samples from post-natal (P) day 5; 72 from P6; 60 from P7; 46 from P8; 18 from P9; 20 from P10; 14 from P12 and 24 from P13. After killing animals by cervical dislocation followed by decapitation, the bulla was removed and opened in Leibowitz medium. The bony capsule of the cochlea was chipped away, stria vascularis and spiral ligament were removed and the sensory epithelium was carefully dissected out of the modiolus. Total RNA was extracted using the micro Fasttrack kit (catalog # K1593-02; Invitrogen, Carlsbad, CA), according to manufacturer's instructions. Reverse transcription and library construction were carried out with the Uni-Zap XR vector kit (catalog # 237211, Stratagene) and Uni-Zap XR Gigapack III Gold Cloning kit (catalog # 237612), both from Stratagene (La Jolla, CA, USA), according to manufacturer's instructions. Briefly: 1.5 ug mRNA was reverse transcribed using a hybrid oligo(dT) linker-primer that contains an Xho I site. First strand synthesis was primed with the linker- primer and transcribed using Moloney murine leukemia virus reverse transcriptase (MMLV-RT) and 5-methyl dCTP. The second strand was synthesized with DNA polymerase and RNase H. Complementary DNA was blunt ended with Pfu DNA polymerase, ligated with EcoR I adapters in the presence of ligase and digested with Xho I. The cDNA was sequentially size fractionated over Pharmacia Size Sep400 (Pharmacia, Uppsala, Sweden) and Clontech Chroma Spin-1000 (Clontech, Palo Alto, CA) columns to enrich for cDNAs greater than 400bp and 1000 bp, respectively. The cDNA was then directionally ligated to the Uni-ZAP XR vector, which had been predigested with EcoR I and Xho I. The phagemid was packaged with Gigapak III Gold and, upon titration on XL1 Blue MRF' cells, the yield of the phage library was estimated to be 11,100,000 recombinants. Stratagene's ExAssist Interference resistance helper phage (catalogue # 211203) was adopted to rescue plasmid DNA from the phages. Upon plating of the rescued library, individual cDNA clones were selected and grown in 96-well, 2 ml growth plate. Plasmid DNA was purified from 200 ul of saturated culture with the Concert96(TM) plasmid purification kit (Invitrogen, Carlsbad, CA) as instructed by the manufacturer. ESTs from the 5' end of the cDNA clones were generated with the universal M13 reverse primer (CAGGAAACAGCTATGACC) and 25% strength BigDye terminator sequencing chemistry (Applied Biosystems, Foster City, CA). Sequencing reactions were performed on MJ Tetrad thermal cyclers (MJ Research, Waltham, MA), and analyzed on 3700 automated capillary sequencers using POP5 polymer (Applied Biosystems, Foster City, CA). The frequency distribution of the library is as follows: 72% of genes have 1 copy; 14.3% 2; 12% 3-10; 1.4% 11-50 and 0.1% 51-150. As to gene function, 45% of genes are present in GenBank and have know function; 23% have hits in GenBank, but do not have assigned function; 12% are uncharacterized ESTs and 20% are unidentified."

ORIGIN

Query Match 6.5%; Score 54.4; DB 13; Length 278;
Best Local Similarity 53.2%; Pred. No. 34;
Matches 115; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

```
Qy        610 ACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAACCTGTGGTCGCCTGCCGTTTT 669
              ||| |  ||     ||     || ||| | | |  |||| |   |||| ||   |
Db         12 ACTGCTCTTCCAAAGGTCCGAAGTTCAAATCCCAGCAACCACATGGTGGCTCACAACCAC 71

Qy        670 TTCAGAGTCAAATCAAGGCACAAATACGCAGTGGCGTAAAAAACGTAGATTCTGATTTTA 729
              |   |    | ||  |   |  | | | || | |  || | |  |   |
Db         72 CCATAACGAGATCTGATGCCCTCTTCTGGTGCGTCTGAAGACATCTACAGTGTACTTACA 131

Qy        730 GCTTATAGAAGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAAACTTTTGAAAAC 789
              | |||  |   |  ||  || || |  | | |   || ||| | ||||  ||| ||||
Db        132 TATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATCTTTAAAAAA 191

Qy        790 AACGCCATAAAAAAAAAAAAAAAAAAAAAAAAAAA 825
              ||    | ||||||||||||||||||||||||||||
Db        192 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 227
```

RESULT 12
BX403488/c
LOCUS BX403488 539 bp mRNA linear EST 15-MAY-2003
DEFINITION BX403488 Homo sapiens PLACENTA Homo sapiens cDNA clone CL0BA002ZE10 3-PRIME, mRNA sequence.
ACCESSION BX403488
VERSION BX403488.1 GI:30762412
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 539)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7316.r
Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CL0BA002ZE10FP1.
FEATURES Location/Qualifiers
source 1. .539
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CL0BA002ZE10"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."
ORIGIN

Query Match 6.5%; Score 54.4; DB 13; Length 539;
Best Local Similarity 55.0%; Pred. No. 25;
Matches 88; Conservative 7; Mismatches 65; Indels 0; Gaps 0;

```
Qy        666 TTTTTTCAGAGTCAAATCAAGGCACAAATACGCAGTGGCGTAAAAAACGTAGATTCTGAT 725
              ||| :| | |   |||  ||   | ||| |   | |     |  |||   |: ::: |  |
Db        301 TTTAWTTAAAAAAAAAAAAAAATATAAAAAAAAAATNNTTTTTAAATTTTWWWWWTTTT 242

Qy        726 TTTAGCTTATAGAAGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAAACTTTTGA 785
              |||   || || ||  |  ||  || ||    ||| ||  | |||  |  ||  ::  |
Db        241 TTTTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAATAAAAWWAAA 182
```

```
Qy        786 AAACAACGCCATAAAAAAAAAAAAAAAAAAAAAAAAAAA 825
              ||| ||    | |||||||||||||||||||||||||||||
Db        181 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 142
```

RESULT 13
CK155159/c
LOCUS CK155159 872 bp mRNA linear EST 05-DEC-2003
DEFINITION FGAS033880 Triticum aestivum FGAS: TaLt2 Triticum aestivum cDNA, mRNA sequence.
ACCESSION CK155159
VERSION CK155159.1 GI:38976978
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 872)
AUTHORS Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D., Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A., Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D., Penniket,C., Roach,J.L. and Sarhan,F.
TITLE Functional Genomics of Abiotic Stress In Wheat and Canola Crops
JOURNAL Unpublished (2003)
COMMENT Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon, Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas_ests@cs.usask.ca
This sequence is the direct result of the Base calling software Phred (default parameters). It is the raw base calls. To aid in the identification of the high quality insert the software Lucy (default parameters) has been run on this sequence. Lucy identified the region [127,285].
Plate: TaLt260 row: D column: 01.
FEATURES Location/Qualifiers
source 1. .872
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Wheat line CI 14106"
/db_xref="taxon:4565"
/lab_host="DH5 alpha"
/clone_lib="Triticum aestivum FGAS: TaLt2"
/note="Organ: Crown; Vector: pGEM-T; SSH (suppression subtractive hybridization) cDNA library from genotype CI14106 cold hardened at 2 C for 1 day (24 H)(tester) and subtracted against genotype Norstar cold hardened at 2 C for 21 days and 49 days (equal amount of cDNA pooled together before subtraction, driver). Nitro-pyrole anchored oligo-dT priming and non-directional cloning."
ORIGIN

Query Match 6.4%; Score 54.2; DB 14; Length 872;
Best Local Similarity 60.5%; Pred. No. 21;
Matches 89; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy        679 AAATCAAGGCACAAATACGCAGTGGCGTAAAAAACGTAGATTCTGATTTTAGCTTATAGA 738
              |||  ||   | |||  |||       |||||||  ||||     ||| | |   | | |
Db        312 AAAAAAAAAAAAAAACCCGCCCCAAAATAAAAAAAATAGAAGTAAATTGTTGAAAAAAAA 253

Qy        739 AGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAAACTTTTGAAAACAACGCCATA 798
              |||  |||  || ||     ||| || ||  ||  |  |||      |||| ||    | |
Db        252 AGTGAGAAAAAAAAAAAAGAAAAAATCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 193

Qy        799 AAAAAAAAAAAAAAAAAAAAAAAAAAA 825
              |||||||||||||||||||||||||||
Db        192 AAAAAAAAAAAAAAAAAAAAAAAAAA 166
```

```
RESULT 14
AL558886
LOCUS       AL558886                 918 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL558886 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
            Homo sapiens cDNA clone CS0DJ015YH06 5-PRIME, mRNA sequence.
ACCESSION   AL558886
VERSION     AL558886.2  GI:31283019
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 918)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12903846.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            6274.f For more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DJ015DD03QP1&cluster=6274.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DJ015DD03QP1.
FEATURES             Location/Qualifiers
     source          1. .918
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ015YH06"
                     /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
                     /cell_line="JURKAT"
                     /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
                     10-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             6.4%;  Score 54.2;  DB 9;  Length 918;
  Best Local Similarity   62.9%;  Pred. No. 20;
  Matches   78;  Conservative    3;  Mismatches   43;  Indels    0;  Gaps    0;

Qy        715 TAGATTCTGATTTTAGCTTATAGAAGTAGGAACGAAGAAGTGTAAACAACCATTAATGAT 774
              ||| |||||| |||  |  |  ||||||  |  ||||||  ||| ||  |  ||  ||
Db        698 TAGTTTCTGAGTTTTACCAAATAAAGTAGAATATAAGAAGAAAAAAAAAAAAAAAAAAAT 757

Qy        775 TAAACTTTTGAAAACAACGCCATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGGCCGCT 834
              |      :   |||| ||    | ||||||||||||||||||||||| ||||:|  || |
Db        758 TTTKGGBNRAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAASCCCCCCCC 817

Qy        835 CGAA 838
              | :|
Db        818 CCMA 821


RESULT 15
CNS00KB5
LOCUS       CNS00KB5                 1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR39P05 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL077453
VERSION     AL077453.1  GI:4956930
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR39P05"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             6.4%;  Score 54.2;  DB 29;  Length 1101;
  Best Local Similarity   56.5%;  Pred. No. 19;
  Matches   83;  Conservative    8;  Mismatches   56;  Indels    0;  Gaps    0;

Qy        679 AAATCAAGGCACAAATACGCAGTGGCGTAAAAAACGTAGATTCTGATTTTAGCTTATAGA 738
              |||  ||   | |||||   | |    |||||||  || |   |  ||  |  |:||| |
Db        138 AAAAAAAAAAAAAAAATAATAAATAATATAAAAAAATTATAAAATATTTAAAWATWATAAA 197

Qy        739 AGTAGGAACGAAGAAGTGTAAACAACCATTAATGATTAAACTTTTGAAAACAACGCCATA 798
              | ||   :     || |  ::| ||  | :||| |  |||  : | |||| |:    |||
Db        198 AATAAATWWATTAAAATTAWWAAAAWAAAWAATTAAAAAAAAAWATAAAAAAAWAAAWATA 257

Qy        799 AAAAAAAAAAAAAAAAAAAAAAAAAAA 825
              :||||||||| ||||||||||||||||
Db        258 WAAAAAAAATAAAAAAAAAAAAAAAA 284


Search completed: August 13, 2004, 20:45:37
Job time : 2786.5 secs
```